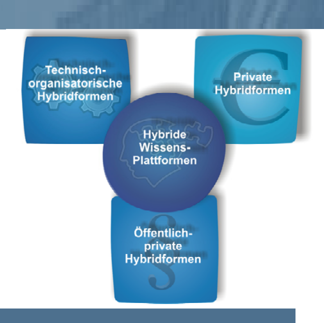


Prof. Dr. Dieter Ahlert, PD Dr. Detlef Aufderheide, Prof. Dr. Klaus Backhaus,
Prof. Dr. Jörg Becker, Prof. Dr. Heinz Lothar Grob, Prof. Dr. Karl-Hans Hartwig,
Prof. Dr. Thomas Hoeren, Prof. Dr. Heinz Holling, Prof. Dr. Bernd Holznagel,
Prof. Dr. Stefan Klein, Prof. Dr. Andreas Pfingsten, Prof. Dr. Klaus Röder.

Internetökonomie und Hybridität

Nr. 51

HEINZ LOTHAR GROB, GOTTFRIED VOSSEN (HRSG.)

Entwicklungen im Web 2.0 aus technischer, ökonomischer und sozialer Sicht



European Research Center
for Information Systems



Westfälische
Wilhelms-Universität
Münster

Gefördert durch:



Förderkennzeichen:
01 AK 704

Projektträger:





Koordination Internetökonomie und Hybridität

Christian Buddendick
Christian.Buddendick@ercis.de
www.ercis.org

NR. 51

Grob, H. L., Vossen, G. (Hrsg.)

Entwicklungen im Web 2.0 aus technischer, ökonomischer und sozialer Sicht

INSTITUT FÜR WIRTSCHAFTSINFORMATIK DER WESTFÄLISCHEN WILHELMS-UNIVERSITÄT MÜNSTER
LEONARDO-CAMPUS 3, 48149 MÜNSTER, TEL. (0251) 83-38000, FAX. (0251) 83-38009

EMAIL: GROB@UNI-MUENSTER.DE

http://www-wi.uni-muenster.de/aw/

Oktober 2007

Vorwort

Web 2.0 ist das wohl bekannteste aktuelle Schlagwort im Umfeld des Internets und des World-Wide Webs. Eine explodierende Zahl von Blogs (z. B. die Web 2.0-Workgroup), Projekten (z. B. Go2Web20.net), Zeitschriften und deren diverse Sonderausgaben (money.cnn.com/ magazines/business2/, www.heise.de/kiosk/special/ix/07/01/), die bekannten „Summit“-Konferenzen von O’Reilly (www.web2summit.com) und Nachrichten haben dabei eine Vielzahl von Definitionen und Beschreibungen dieses Begriffs hervorgebracht. Er fasst verschiedene Trends der letzten Jahre aus zunächst unabhängig erscheinenden Entwicklungen zusammen und betrifft sowohl Fragen der Nutzung des Internets, seiner potenziellen Anwendungsmöglichkeiten und Auswirkungen auf die Gesellschaft als auch der technischen Umsetzung. Während zahlreiche Quellen sich allerdings darauf beschränken, die Phänomene zu erklären, sind Hunderte von Start-up-Firmen vor allem im Silicon Valley dabei, diese in Anwendungen umzusetzen, mit welchen sich direkt oder indirekt Geld verdienen lässt.

Bei allen diesen Aktivitäten und Entwicklungen, die in erster Linie (wieder einmal) technologisch getrieben sind, ist es an der Zeit, sich auch aus Forschungssicht mit dem Web 2.0 auseinanderzusetzen. Wir tun dies am Institut für Wirtschaftsinformatik der Westfälischen Wilhelms-Universität Münster seit dem Sommer 2006 im Rahmen von Seminaren und Forschungsprojekten und haben dies im Sommer 2007 erstmals auch im Rahmen eines Doktorandenseminars getan. Die vorliegende Ausarbeitung ist das Ergebnis dieses Seminars. Insgesamt 19 Teilnehmer haben sich mit Entwicklungen im Web 2.0 aus technischer, ökonomischer und sozialer Sicht beschäftigt, und zwar mit der Maßgabe, dass sich die aktuellen Entwicklungen in vier größere Bereiche einteilen lassen, die sich gegenseitig beeinflussen:

1. Benutzererzeugte Inhalte, was den Übergang vom „Read-Web“ zum „Read/Write-Web“ markiert;
2. Geschäftliche Aspekte, welche sich mit der Frage nach Geschäftsmodellen und der Monetisierbarkeit des Webs befassen;
3. Suchen und Sozialisierung, was inzwischen neue Dimensionen erreicht – u. a. dadurch, dass Software und Dienste im Netz durch mehr Benutzer besser werden;
4. technische Aspekte und Infrastrukturen, welche heute Kapazitäten und Geschwindigkeiten erst ermöglichen.

Wir danken allen Teilnehmern des Seminars für ihre Beiträge, die einen guten Überblick über aktuelle Entwicklungen und deren Perspektiven aufzeigen, und wüschen ihnen, dass das hier Erfahrene auch in das eigene Dissertationsprojekt einfließen wird. Unseren Lesern wünschen wir neue Erkenntnisse beim Stöbern in diesem Bericht.

Münster, im Oktober 2007

Heinz Lothar Grob, Gottfried Vossen

Inhalt

I User-Generated Content

Tagging vs. Ontologies

David Stadler

1 Einleitung

Im Zuge der Entwicklung des Internets und der sich in diesem Zusammenhang wandelnden Rolle des Nutzers haben sich seit der Jahrtausendwende neue Arten zur Beschreibung von Inhalten im Internet entwickelt. Während in den Anfangszeiten des Internets der Nutzer fast ausschließlich die Rolle des Konsumenten einnahm, hat sich diese mit neuen Technologien und der Verbreitung schneller Internetanschlüsse hin zu dem so genannten Prosumer verändert. Hinter dieser Wortschöpfung verbirgt sich eine Kombination aus Konsument und Produzent, da die Nutzer nicht mehr nur noch Inhalte konsumieren, sondern auch zunehmend selber produzieren. Die damit einhergehende große Herausforderung ist es, die Informationsflut im Internet in Form von Blogs, Fotos etc. auf eine nutzerfreundliche Art zu ordnen.

Im Kapitel 2 des vorliegenden Beitrags werden zwei Arten zur Einordnung und Beschreibung von Inhalten im Internet vorgestellt und verglichen. Zum einen sind es Ontologien, die ein vorgegebenes Vokabular verwenden und definierte Beziehungen zwischen den Beschreibungsmerkmalen beinhalten. Zum anderen werden Folksonomien beleuchtet, die mithilfe von Tags (frei wählbaren Schlagworten) eine Beschreibung von Inhalten erlauben, ohne dass dabei vorgegebene Konventionen eingehalten werden müssen. Kapitel 3 widmet sich verschiedenen Klassifizierungsmerkmalen von Folksonomien. In Kapitel 4 werden diverse Probleme aufgezeigt, die mit Folksonomien und der freien Vergabe von Tags einhergehen und bei einem kontrollierten Vokabular, wie es bei Ontologien der Fall ist, kaum präsent sind. Dies sind etwa Probleme mit Homonymen und Synonymen, der Verwendung unterschiedlicher Sprachen oder dem Einsatz von Sonderzeichen. Trotz dieser Probleme ist der Einsatz von Tagging inzwischen weit verbreitet. Daher wird in Kapitel 5 untersucht, warum Folksonomien funktionieren und auch äußerst erfolgreich sind. Schließlich werden im letzten Kapitel einige Anregungen gegeben, wie man Ontologien und Folksonomien gemeinsam einsetzen kann, um auf diese Weise von den Vorteilen beider Arten, Inhalte zu beschreiben und zu klassifizieren, profitieren zu können.

2 Systeme zur Beschreibung von Inhalten

2.1 Ontologien

Zur Beschreibung und Einordnung von Inhalten im Internet werden Metadaten verwendet, welche auch als Daten über Daten [Ber97] bezeichnet werden. Als Inhalte kommen dabei Bilder, Videos, Artikel und sämtliche mittels einer URL eindeutig identifizierbaren Objekte infrage. Beschreibende Metadaten unterstützen User, aber auch Agenten (intelligente Softwareprogramme) bei der Identifikation von relevanten gegenüber irrelevanten Inhalten [MM06].

Damit sich Metadaten gemeinsam verwenden lassen und sich zum Austausch über bestimmte Themenbereiche eignen, müssen sie einem einheitlichen Format genügen. Traditionell werden

Metadaten zur Kategorisierung von Inhalten daher professionell erzeugt. Beispielhaft sei an dieser Stelle auf die Klassifikationen in den Bibliothekswissenschaften verwiesen, in denen bisher umfangreiche Schemata und Regeln entstanden sind, wie etwa die Dewey Decimal Classification (DCC). Kennzeichnend für solche Systeme ist die eindeutige Klassifizierung aller Inhalte, ohne dabei zweideutige Einordnungsmöglichkeiten zuzulassen. Die Kategorien, in denen die Inhalte eingeordnet werden, sind ihrerseits wiederum in höheren Kategorien enthalten und bilden mithin eine Hierarchie. Werden zusätzlich zu der einfachen Klassifizierung durch Metadaten noch Beziehungen zwischen den einzelnen Inhalten abgebildet, spricht man auch von einer Ontologie.

Ursprünglich kommt der Ontologiebegriff aus der Philosophie. Eine Ontologie im philosophischen Sinne hat das Ziel, die Realität mittels einer Klassifikation aller Entitäten zu beschreiben. Im Sinne der Informatik definiert GRUBER eine Ontologie als „an explicit specification of a conceptualization“ [Gru93][Fen01]. Eine Konzeptualisierung meint eine vereinfachte Darstellung aller relevanten Objekte und deren Beziehungen untereinander. Folglich eignet sich eine Ontologie dazu, komplexe Strukturen (spezieller Wissensbereiche) abzubilden und ein gemeinsames Verständnis von Begriffen und Relationen zwischen diesen herbeizuführen [Fen00].

Gemäß der Idee des Semantic Webs sollen Ontologien dabei helfen, Inhalte durch Anreicherung mit semantischen Metadaten maschinenlesbar zu machen und somit den Wissensaustausch zwischen Maschinen und zwischen Mensch und Maschine zu ermöglichen [BHL01]. Als Konzept in diesem Bereich, Inhalte sowohl mit Metadaten zu versehen als auch die Beziehungen zwischen diesen abzubilden, ist Web Ontology Language (OWL) [MH04] weit verbreitet, welches wiederum eine Erweiterung von Resource Description Framework (RDF) [LS99] darstellt. Ein einfacher RDF-Graf ist ein Tripel, bestehend aus einem Subjekt, einem Prädikat und einem Objekt.

Ontologien entwickeln sich mit neuen wissenschaftlichen Erkenntnissen und daraus resultierendem Wissen hinsichtlich der Domäne einer Ontologie weiter, sodass diese angepasst werden muss [Mik05][Shi05]. Hier zeigt sich eine wesentliche Schwierigkeit der Ontologien. Das abgebildete Wissen ist nie abschließend, daher kann eine Ontologie dem Anspruch, sämtliches Wissen eines Themenbereiches zu kategorisieren, kaum gerecht werden.

2.2 Tagging

Ein entscheidender Nachteil der professionellen Erstellung von Metadaten besteht darin, dass die Erzeugung sehr aufwendig, damit teuer und gerade bei sehr großen Datenmengen nicht beliebig skalierbar ist [DHS+02][Mat04]. Um diesem Nachteil entgegenzuwirken, gibt es verschiedene Ansätze, wer die Annotation der Inhalte mit Metadaten vornehmen soll. Eine Alternative zur professionellen Erzeugung von Metadaten stellen von Autoren annotierte Inhalte dar. Mithilfe von Autoren kommt man somit dem Problem der Skalierung bei. Allerdings müssen die Autoren hierfür die Konventionen der Erstellung von Metadaten kennen und sich in die entsprechende Ontologie zunächst einarbeiten. Problematisch ist ferner, dass die Metadaten nicht zwangsläufig den Anforderungen der Nutzer genügen, Inhalte (wieder)zufinden oder in Wissensgebieten zu recherchieren [Mat04]. Von Nutzern erstellte Metadaten bilden daher einen weiteren Ansatz, Inhalte mit Metadaten anzureichern.

Auf diese Weise erzeugte Schlagworte zu beliebigen Objekten werden als Tagging bezeichnet. Schlagworte können dabei frei wählbare Begriffe sein, welche auch Tags genannt werden. Ziel des Taggings ist es, die mit den Tags verknüpften Objekte besser wiederfinden oder durchsuchen zu können [GH06]. Da die Nutzer bei der Erzeugung der Tags nicht an bestimmte Muster gebunden sind, können diese beliebige Assoziationen mit dem Objekt mittels seiner Tags ausdrücken, welche für Außenstehende nicht unbedingt Sinn ergeben müssen [GT06]. Das Tagging wird von den Nutzern als besonders einfach gegenüber dem Einordnen in eine Ontologie empfunden, da es dem kognitiven Denken des Menschen am nächsten kommt, ohne dass eine Entscheidung getroffen werden muss, wie das Objekt eindeutig einzuordnen ist [Sin05]. So fällt es dem menschlichen Gehirn leichter, einzelne Worte im Gegensatz zu ausformulierten inhaltlichen Beschreibungen oder Kommentaren zu erfassen.

Inhaltlich lassen sich Tags in verschiedene Kategorien einordnen, die etwas über die annotierte Ressource aussagen. Bei der folgenden Einteilung gängiger Tags wird deutlich, dass Tags nicht nur zur inhaltlichen Beschreibung verwendet werden [GH06]:

- Identifying what (or who) it is about: Diese Art von Tags kommt am häufigsten vor und bezieht sich ausschließlich auf das Themengebiet der Ressource. Hierfür werden i. d. R. Nomen verwendet.
- Identifying what it is: Bei dieser Gruppe von Tags handelt es sich um eine Beschreibung der Art der Ressource, losgelöst von ihrer inhaltlichen Beschreibung. Beispielhaft seien an dieser Stelle Tags genannt wie Blog, Article und Book.
- Identifying who owns it: In diese Kategorie fallen Tags, die eine Aussage darüber treffen, wem die getaggte Ressource gehört, wie etwa der Autor eines Blogs.
- Refining Categories: Diese Gruppe kann nie unabhängig von anderen Tags verstanden werden, da diese zur weiteren Verfeinerung bestehender Tags dient.
- Identifying Qualities or Characteristics: Tags dieser Art beschreiben den Inhalt i. d. R. mithilfe passender Adjektive. Auf diese Weise kann der Nutzer seine Meinung über eine Ressource artikulieren, indem er sie als funny, stupid etc. bezeichnet.
- Self Reference: Hierunter fallen Tags, die eine Beziehung des Nutzers zur Ressource ausmachen, wie etwa *mystaff* oder *mycomments*.
- Task Organizing: Bestandteil dieser Kategorie sind schließlich Tags, die zur Organisation verwendet werden, wie beispielsweise *toread* oder *jobsearch*.

2.3 Folksonomien (Collaborative Tagging)

Wenn viele Nutzer innerhalb eines Systems gemeinsam Tags erzeugen und diese Tags auch anderen Nutzern zugänglich machen, spricht man von einer Folksonomie. Der Begriff folksonomy wurde erstmals 2004 von VANDER WAL geprägt und setzt sich zusammen aus folk und taxonomy [Smi04].[1] Da es sich bei einer Taxonomie allerdings um eine hierarchische Liste von kon-

---

[1] Erstmalig veröffentlicht wurde der Begriff in einem Weblog von SMITH im Nachgang einer Diskussion mit den Mitgliedern des Information Architecture Institute, denen auch VANDER WAL angehörte.

trolliertem Vokabular handelt, ist der Begriff nicht ganz passend gewählt [Mat04] [Mer04]. Dennoch wird im weiteren Verlauf der Arbeit der Begriff Folksonomie weiterverwendet.

Mathematisch lässt sich eine Folksonomie als ein tripartiter Graph darstellen [Mik05]. Es gibt drei Mengen von Knoten, nämlich die Menge von Nutzern, Tags und Ressourcen. Die weiteren Ausführungen beschäftigen sich insbesondere mit Teilgrafen aus einer Folksonomie, nämlich der Relation zwischen Nutzern und Tags, zwischen Tags und Ressourcen und zwischen Nutzern und Ressourcen. Die erste Relation wird in Kapitel 5 genauer beleuchtet, während in Kapitel 3 auf einzelne Aspekte der beiden nachfolgenden Beziehungen eingegangen wird. Innerhalb einer Folksonomie ergeben sich allerdings auch Verbindungen zwischen den Nutzern mittels der Kanten, indem mehrere Nutzer sich für das gleiche Themengebiet interessieren oder sich durch Tags beeinflussen und sogar eine gemeinsame Art zu taggen entwickeln.

Der Übersicht halber werden Tags in einer Folksonomie oftmals in einer sog. Tag Cloud dargestellt. Diese zeigt ein Subset (meist die 100 häufigsten) von Tags zu einer Ressource oder zu allen Ressourcen in meist alphabetischer Reihenfolge an und hebt besonders häufig verwendete Tags durch eine größere Schriftart hervor.[2] Erstmals wurden Tag Clouds 2004 bei Flickr [1] eingeführt, sind aber mittlerweile bei fast allen Folksonomie-Systemen fester Bestandteil.

3 Differenzierung von Folksonomien

Das Grundprinzip von Folksonomien lässt sich in zahlreichen Anwendungen im Internet wiederfinden. Der Erfolg dieser Seiten liegt darin, dass kein spezielles Wissen benötigt wird und die User einen direkten Nutzen von den Anwendungen haben [HJS+06]. Grundsätzlich lassen sich diese Systeme nach sieben Kriterien unterscheiden [MNB+06]:

- Tagging Rights: Hier stellt sich die Frage, wer die Ressourcen taggen darf. Das Spektrum reicht dabei von Systemen, die nur das Taggen von eigenen Ressourcen ermöglichen (z. B. Technorati), bis hin zu offenen Systemen, die jedem Nutzer das Taggen aller Ressourcen erlauben (z. B. Del.icio.us [2]). Zwischen diesen beiden Extremen kann man allerdings noch beliebig viele Mischformen finden, wie etwa bei Flickr eine definierte Gruppe von Freunden, die ebenfalls die eigenen Bilder taggen darf. Ebenso wie bei der Zuordnung von Tags gibt es ein ähnlich weites Spektrum unterschiedlicher Rechte beim Ändern oder Löschen von Tags. VANDER WAL verwendet eine ähnliche Einteilung der Systeme in breite und enge Folksonomien [Van05]. Breite Folksonomien erlauben im Gegensatz zu engen Folksonomien jedem User das Taggen aller Ressourcen.
- Tagging Support: Tagging Support bezieht sich auf die Hilfen, die dem Nutzer beim Taggen angeboten werden. Blind Tagging findet dann statt, wenn der Nutzer bei der Vergabe von eigenen Tags nicht auf bereits angelegte Tags zurückgreifen kann. Beim Viewable Tagging werden dem Nutzer die bereits bestehenden Schlagworte zu einer Ressource angezeigt, während beim Suggestive Tagging sogar Tags vorgeschlagen werden.

---

[2] Siehe als Beispiel eine Tag Cloud bei Flickr [1] (http://www.flickr.com/photos/tags/) oder del.icio.us [2] (http://del.icio.us/tag/). Weitere Informationen zu Tag Clouds finden sich beispielsweise bei [HH06].

- Aggregation: Einen weiteren Gesichtspunkt stellen Aggregationen von Tags dar. Während manche Systeme zu einer Ressource beliebige Tag-Gruppen zulassen und somit auch sich wiederholende Tags nicht verbieten (z. B. Del.icio.us), verhindern andere Systeme dieses, indem Tags zusammengefasst werden bzw. sich die Gruppe der Nutzer auf verschiedene Tags einigt (s. YouTube, Flickr). Beliebige Tag-Gruppen zu einer Ressource werden nach MARLOW U. A. als bag-model bezeichnet, während die andere Form als set-model deklariert wird. Bei einem bag-model sollte das System Möglichkeiten anbieten, die vergebenen Tags zu analysieren, um auf diese Weise häufige Tags zu identifizieren und Relationen zwischen Nutzern, Ressourcen und vergebenen Tags aufzudecken.
- Type of Object: Den Anfang getaggter Ressourcen bildeten Webseiten. Die Liste getaggter Ressourcen wurde aber im Laufe der Zeit schnell um beispielsweise Bilder, Videos, Blogs etc. ergänzt. Grundsätzlich kann jede Ressource getaggt werden, die sich virtuell abbilden lässt. Der Kreis getaggter Ressourcen lässt sich somit beliebig um sämtliche Ressourcen erweitern, die sich eindeutig durch eine URL identifizieren lassen. Mit den verschiedenen Ressourcen, die in den Systemen getaggt werden, ergeben sich laut MARLOW U. A. auch unterschiedliche Arten von Tags. So werden beispielsweise Webseiten anders getaggt als es bei Bildern, Videos oder Musik der Fall ist.
- Source of Material: Die Ressourcen in Folksonomien können aus unterschiedlichen Quellen stammen. Das Herkunftsspektrum reicht dabei von den Nutzern über das System selber bis hin zu allen Quellen, die das Internet bietet. Dem Inhalt und der Philosophie der Plattformen entsprechend wird allerdings die Herkunft beschränkt. Bei Flickr können die Ressourcen (Bilder) nur vom Fotografen selber oder vom Besitzer der Bilder kommen.
- Resource Connectivity: Die Links zwischen den Ressourcen lassen sich in drei Kategorien zusammenfassen: linked, grouped und none. Linked sind beispielsweise Internetseiten, welche über Links direkt miteinander verknüpft sind. Grouped sind bei Flickr Fotos, die sich durch Zeit, Ort oder Tätigkeit zusammenfassen lassen, an oder zu denen die Bilder entstanden sind. Ressourcen, die keinerlei Verbindung aufweisen, fallen in die letzte Kategorie.
- Social Connectivity: Entsprechend der Verbindung der Ressourcen lassen sich auch die Verbindungen der Nutzer einteilen. So erlauben manche Systeme die Verbindung von Nutzern zu Gruppen. Bei Flickr besteht beispielsweise die Möglichkeit, Freunde zu definieren.

4 Probleme und Möglichkeiten von Folksonomien

4.1 Homonyme und Synonyme

Mit der Nutzung von beliebigen Wörtern als Tags gehen einige potenzielle Probleme einher, mit denen die Systeme, die Tags einsetzen, umgehen müssen. In Folksonomien werden beispielsweise keine Synonyme gefiltert [Mat04]. Tags mit gleicher Bedeutung, wie beispielsweise „Mac“, „Macintosh“ oder „Apple“, kommen häufig vor. Eine weitere Herausforderung stellen Homonyme[3] dar. Bei Homonymen handelt es sich um Wörter, welche mehrere Bedeutungen ha-

---

[3] Im Kontext dieser Arbeit werden mehrdeutige Wörter (Äquivokationen) unter dem Begriff Homonym zusammengefasst. Eine weitere Untergliederung nach Homonymen und Polysemen soll nicht stattfinden.

ben, wie etwa „Jaguar“ oder „Apple“. Gleichzeitig verursachen unterschiedliche Formen von Wörtern beispielsweise im Plural oder Singular ein verwandtes Problem.

Die Systeme versuchen, mit unterschiedlichen Mechanismen diesen Problemen beizukommen. So gibt es etwa bei Del.icio.us die Möglichkeit, sich verbundene Tags (Related Tags) anzeigen zu lassen. Hierüber ist es i. d. R. möglich, einen Zusammenhang zwischen den Tags herzustellen und somit die intendierte Bedeutung eines Homonyms herauszufinden [GH06]. Bei Synonymen hingegen gestaltet sich das Problem ungleich schwieriger, da das System in der Lage sein müsste, sämtliche Tags herauszufiltern, welche dieselbe Ressource identifizieren. In einer Folksonomie besteht allerdings die Möglichkeit, dass sich die Nutzer auf die Verwendung von einem oder einigen wenigen Begriffen einigen und somit bei der Suche nach bestimmten Ressourcen nicht erst alle synonym verwendeten Tags herausfinden müssen [GH06]. Ähnliche Ansätze werden bei der Verwendung von Plural oder Singular verfolgt, wobei die Nutzer sich auf bestimmte Konventionen einigen, allerdings ohne dass deren Einhaltung überwacht wird.

Ähnliche Probleme sind in Ontologien kaum präsent, da hier ein kontrolliertes Vokabular sicherstellt, dass gewisse Konventionen eingehalten werden – beispielsweise bei der Verwendung von Plural oder Singular. Gleichzeitig verhindern die hierarchischen Kategorien, dass etwa ein Jaguar fälschlicherweise als eine Automarke interpretiert wird.

4.2 Leerzeichen und Sonderzeiten

Die meisten Systeme erlauben als Tags nur einzelne Begriffe. Daher bedienen sich die Nutzer unterschiedlicher Trennzeichen, um zusammengesetzte Tags zu erzeugen. Untersuchungen zufolge wird hierbei meist der einfache Bindestrich verwendet [GT06]. Aber auch andere Zeichen (Slash, Unterstrich etc.) werden nicht selten eingesetzt. Außerdem finden Sonderzeichen Verwendung, um Tags bei einer alphabetischen Sortierung vorne zu platzieren sowie um eine weitere Semantik zu unterstützen. So lassen sich die Arten von Tags auch mithilfe von Sonderzeichen darstellen, welche die Semantik unterstreichen. Organisatorische Angaben werden z. B. durch ein Ausrufungszeichen (!toread) oder die Bewertung einer Ressource durch eine vorangestellte Tilde (~funny) bzw. durch die Vergabe von Sternchen für besonders gute Quellen (*****) dargestellt.

Entsprechende Sonderzeichen können verwendet werden, um ein einheitliches Tagging in Folksonomien zu erzielen. Allerdings geht eine solche Vorgabe der Syntaktik mit einer Einschränkung bei der Vergabe von Tags einher. So könnte das System beispielsweise ähnlich einer Ontologie hierarchische Strukturen aufbauen, indem durch Schrägstriche getrennte Tags ausgewertet werden, die eine Folge spezialisierender Tags widerspiegeln.

4.3 Sprachen

Neben Sonderzeichen werden in Folksonomien aufgrund der unterschiedlichen Herkunft der Nutzer auch verschiedene Sprachen verwendet. Dabei überwiegt aber mit großem Abstand die

englische Sprache weit vor anderen Sprachen wie etwa Spanisch.[4] Dennoch verursachen unterschiedliche Sprachen weitere Probleme, die sich nicht nur in einer Sprachbarriere für die Nutzer äußern. Insbesondere Begriffe, die in unterschiedlichen Sprachen verschiedene Bedeutungen haben, können zu Irritationen bei Nutzern aus unterschiedlichen Ländern oder Sprachräumen führen.

Indem Regeln geschaffen werden, kann man vielen der genannten Probleme beikommen. So kann es Vereinbarungen geben, als Tags nur englische Begriffe, den Singular und die Kleinschreibung zu verwenden. Wie bereits angesprochen, werden die Nutzer dadurch allerdings entgegen dem Charakter des Taggings in ihrer freien Auswahl beschränkt.

5 Funktionsweise von Folksonomien

5.1 Geringe Eintrittsbarrieren

Eine wesentliche Begründung für die Bildung von Folksonomien liegt im Konzept und in der Architektur begründet. Die Nutzer benötigen keinerlei Vorwissen und können ohne jegliche Einarbeitung Teil einer Folksonomie werden. Laut einem der Gründer von Flickr ist gerade die Vergabe von frei wählbaren Schlagwörtern viel einfacher als die Entscheidung über die Einordnung in eine vorgegebene Kategorie, insbesondere wenn es sich um eine hierarchische Kategorisierung handelt [But04]. Schließlich ist es auch nahezu unmöglich, die große Anzahl der Nutzer einer Folksonomie zur Verwendung eines kontrollierten Vokabulars innerhalb einer Ontologie zu bewegen, da der Zeiteinsatz gegenüber dem Nutzen viel zu hoch wäre [Mat04].

5.2 Power Law

Ein weiterer Grund für den Erfolg von Folksonomien liegt darin, dass sich die vergebenen Tags auf einige Wenige konzentrieren, die immer wieder verwendet werden. Die Verteilung folgt somit dem Power Law, d. h., näherungsweise lassen sich mit 20 % der Tags 80 % der Ressourcen beschreiben. Diese Verteilung wurde bereits von MATHES vermutet [Mat04] und in Untersuchungen am Beispiel von Del.icio.us und Flickr bestätigt [GH06][MNB+06]. Demnach tauchen einige Tags immer wieder auf, während andere nur von wenigen oder sogar nur von einem Nutzer angewendet werden (Long-Tail).

Ein ähnliches Bild ergibt sich auch bei der Untersuchung der Tags eines Nutzers. Während anfangs die Menge der Tags mit der Anzahl getaggter Ressourcen wächst, stabilisiert sich diese Menge, ohne dass mit der Hinzunahme neuer Ressourcen auch in wesentlichem Maße neue Tags herangezogen werden [MNB+06]. Folglich konvergiert die Menge der Tags in Richtung einer definierten Menge von Tags, welche immer wieder benutzt werden.

Untersucht man eine ganze Folksonomie unter Berücksichtigung dieser Ergebnisse über einen längeren Zeitraum, so ergibt sich i. d. R. immer wieder das gleiche Phänomen. Während die Anzahl der Tags zunächst linear wächst, flacht die Kurve, die die Menge verwendeter Tags für

---

[4] Nach einer Untersuchung von GUY, TONKIN überwiegt Englisch mit 65 % bei del.icio.us (45 % bei Flickr), gefolgt von Spanisch mit lediglich 2 %.

eine Ressource darstellt, mit der Zunahme der Tag-Prozesse im Zeitverlauf ab. Die Nutzer der Folksonomie entwickeln demnach mit der Zeit ein gemeinsames Verständnis über eine Ressource, sodass auch immer wieder dieselben Tags verwendet werden. Bei Del.icio.us wird der Effekt noch dadurch verstärkt, dass ein Nutzer die am häufigsten verwendeten Tags vorgeschlagen bekommt, wenn dieser eine neue Ressource annotiert. Allerdings erklärt dieses Vorschlagsystem nicht die komplette Konvergenz von Tags, da auch die nicht so häufig benutzten Tags immer wieder zur Beschreibung herangezogen werden, obwohl diese dem Nutzer nicht angezeigt werden [GH06].

Es steht somit zu vermuten, dass sich die User gegenseitig beeinflussen und Tagging-Angewohnheiten der Community selber aufgreifen. SISHA begründet dieses Phänomen mit dem Instant Feedback, d. h. einer prompten Reaktion auf die vergebenen Tags in Form von Feedback der anderen Nutzer. Dieses Feedback kann sich darin äußern, dass sich Nutzer die Art zu taggen abschauen oder aber die verwendeten Tags infrage stellen. In einer weiteren Studie haben MARLOW U. A. gezeigt, dass eine Gruppe zu einem gemeinsamen Verständnis von Tags tendiert. Hierfür haben sie die paarweise gemeinsam verwendeten Tags eines Nutzers mit einem Kontakt bzw. einem zufällig ausgewählten anderen Nutzer verglichen und festgestellt, dass die Wahrscheinlichkeit für gemeinsame Tags innerhalb der Gruppe größer ist als außerhalb der Gruppe. Folglich beeinflussen sich die Nutzer gegenseitig und scheinen teilweise ein gemeinsames Vokabular zu entwickeln, welches die gesamte Gruppe für das Taggen von Ressourcen verwendet.

6 Konvergenz von Ontologien und Folksonomien

Während die meisten wissenschaftlichen Beiträge eindeutig die Vorteile einer der beiden Seiten (Ontologien oder Folksonomien) hervorheben, schlägt GRUBER vor, beide Ansätze zu verbinden. Seiner Meinung nach ist es nicht immer möglich, eine Ressource in eine vorgegebene hierarchische Systematisierung einzuordnen. Folksonomien hingegen sind einstufig aufgebaut und es gibt keine falschen Entscheidungen bzgl. der Einordnung [Gru05]. So sind Suchmaschinen klar im Vorteil, wenn die zu suchenden Inhalte durch eine Folksonomie getaggt worden sind. Insbesondere bei nicht textbasierten Dokumenten können die Tags Aufschluss über den Inhalt der Ressourcen liefern. Darüber hinaus verhindert eine Auswertung der Tags, dass dem Nutzer zu viele irrelevante Ressourcen als Suchergebnisse präsentiert werden. Allerdings fehlt Folksonomien eine einheitliche Struktur, sodass systemübergreifende Anwendungen nicht auf Tags aus verschiedenen Systemen und zu unterschiedlichen Ressourcen zugreifen können. GRUBER postuliert daher eine Ontologie für Folksonomien [Gru05]. Für einen systemübergreifenden Austausch bedarf es allerdings Konventionen, die von allen Systemen eingehalten werden müssen. Durch Kombination der Tags ist etwa die Beantwortung einer Aufgabe möglich wie: Finde alle Hotels in Spanien, welche mit *romantic* getaggt wurden.

Ein Beispiel für die Kombination von Tags aus unterschiedlichen Kontexten bietet RealTravel [3]. Das System bietet den Nutzern die Möglichkeit, Reiseberichte in Form von Blogs zu schreiben oder Reisepläne inklusive der geplanten Routen zu hinterlegen. Kombiniert werden diese unstrukturierten Inhalte mit konkreten Koordinaten, mit denen sich die Inhalte verknüpfen lassen – allerdings nicht in Form von Tags, sondern durch Auswahl aus entsprechenden Auswahlboxen. Die Inhalte lassen sich zusätzlich mit Fotos anreichern, welche wiederum mit dem ausge-

wählten Ort verbunden sind. Neben der strukturierten Einordnung der Inhalte können diese des Weiteren mithilfe von frei wählbaren Tags beschrieben werden. Dadurch besteht die Möglichkeit, aus der gesamten Folksonomie Nutzer mit ähnlichen (Reise)Interessen herauszufinden und sich auszutauschen.

Einen ähnlichen Ansatz verfolgt Flickr, indem die Fotos mit Kartenmaterial von Yahoo! Maps eingeordnet werden. Neben der Möglichkeit, diese Verknüpfung mittels Auswahl aus einer Weltkarte und beliebigem Zoom zu schaffen, erlaubt Flickr, die Eingabe von Koordinaten auf eine genormte Art und Weise zu taggen. Damit ein Foto korrekt mit seinen Koordinaten verknüpft wird, muss ein Foto mit den folgenden Tags versehen sein:

geotagged, *geo-lat*=<<Breitengrad>>, *geo-long* oder *geo-lon*=<<Längengrad>>

Im Rahmen dieses Kapitels wurden einige Anregungen aufgezeigt, wie Folksonomien und Ontologien verschmelzen können oder gemeinsam genutzt werden können. Tags werden jedoch nie den formalen Grad einer Ontologie erreichen können. Dennoch bietet eine Kombination aus frei wählbaren Tags – eingeordnet in eine strukturierte Ontologie – die Möglichkeit, strukturierte und unstrukturierte Metadaten zu erzeugen, ohne die technischen Fähigkeiten der Nutzer überzustrapazieren.

Literaturverzeichnis

[Ber97] Berners-Lee, T. (1997): Metadata Architecture, http://www.w3.org/DesignIssues/Metadata.html. Zuletzt abgerufen am 16.8.2007.

[BHL01] Berners-Lee, T.; Hendler, J.; Lassila, O. (2001): The Semantic Web. Scientific American, Vol. 284, No. 5, pp. 34-43.

[But04] Butterfield, S. (2004): Sylloge, http://www.sylloge.com/personal/2004/08/folksonomy-social-classification-great.html. Zuletzt abgerufen am 26.8.2007.

[DHS+02] Dumas, E.; Hodgins, W.; Sutton, S.; Weibel, S. (2002): Metadata Principles and Practicalities. D-Lib Magazine, Vol. 8, No. 4. http://www.dlib.org/dlib/april02/weibel/04weibel.html. Zuletzt abgerufen am 18.8.2007.

[Fen00] Fensel, D. (2000): Relating Ontology Languages and Web Standards, in J. Ebert, U. Frank (ed.), Modelle und Modellierungssprachen in Informatik und Wirtschaftsinformatik Koblenz, pp. 111-128.

[Fen01] Fensel, D. (2001): A Silver Bullet for Knowledge Management and Electronic Commerce. Berlin u. a.

[GH06] Golder, S. A.; Huberman, B. A. (2006): The structure of Collaborative Tagging Systems. Journal of information science, Vol. 32, No. 2, pp. 198-208.

[Gru93] Gruber, T. R. (1993): A Translation Approach to Portable Ontology Specification. Knowledge Acquisition, Vol. 5, No. 2, pp. 199-220.

[Gru05] Gruber, T. R. (2005): Ontology of folksonomy. A Mash-up of Apples and Oranges. Proc. of the MTSR. http://tomgruber.org/writing/ontology-of-folksonomy.htm. Zuletzt abgerufen am 25.8.2007.

[GT06] Guy, M.; Tonkin, E. (2006): Folksonomies. Tidying up Tags? D-Lib Magazine, Vol. 12, No. 1. http://www.dlib.org/dlib/january06/guy/01guy.html. Zuletzt abgerufen am 18.8.2007.

[HH06] Hassan-Montero, Y.; Herrero-Solana, V. (2006): Improving Tag-Clouds as Visual Information Retrieval Interfaces. Proc. of the InSciT, Mérida, Spain.

[HJS+06] Hotho, A.; Jäschke, R.; Schmitz, C.; Stumme, G. (2006): Information Retrieval in Folksonomies. Search and Ranking. Proc. of the ESWC. Berlin, Heidelberg, pp. 411-426.

[LS99] Lassila, O.; Swick, R. R. (1999): Resource Description Framework (RDF) Model and Syntax Specification, http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/. Zuletzt abgerufen am 17.08.2007.

[MM06] MacGregor, G.; McCulloch, E. (2006): Collaborative Tagging as a Knowledge Organisation and Resource Discorvery Tool. Library Review, Vol. 55, No. 5, pp. 291-300.

[MNB+06] Marlow, C.; Naarman, M.; Boyd, D.; Davis, M. (2006): HT06, Tagging paper, Taxonomy, Flickr, Academic Article, Toread. Proc. of the Hypertext an Hypermedia. New York, pp. 31-40.

[Mat04] Mathes, A. (2004): Folksonomies – Cooperative Classification and Communication Through Shared Metadata, http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf. Zuletzt abgerufen am 18.8.2007.

[MH04] McGuinness, D. L.; van Harmelen, F. (2004): OWL Web Ontology Language Overview, http://www.w3.org/TR/owl-features/. Zuletzt abgerufen am 17.8.2007.

[Mer04] Merholz, P. (2004): Metadata for the Masses. Adaptive path essays, http://www.adaptivepath.com/publications/essays/archives/000361.php. Zuletzt abgerufen am 19.8.2007.

[Mik05] Mika, P.: Ontologies are us (2005): A Unified Model of Social Networks and Semantics. Proc. of the ISWC. Berlin, Heidelberg, pp. 522-536.

[Shi05] Shirky, C. (2005): Ontology is Overrated. Categories, Links and Tags. Clay Shirky's webblog, http://shirky.com/writings/ontology_overrated.html. Zuletzt abgerufen am 18.8.2007.

[Sin05] Sinha, R. (2004): A Cognitive Analysis of Tagging. Rashmi Sinha's webblog, http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html. Zuletzt abgerufen am 18.8.2007.

[Smi04] Smith, G. (2004): Folksonomy. Social Classification. Gene Smith's webblog, http://atomiq.org/archives/2004/08/folksonomy_social_classification.html. Zuletzt abgerufen am 19.8.2007.

[Van05] Vander Wal, A. (2005): Explaining and Showing Broad and Narrow Folksonomies. Vander Wal's webblog, http://www.vanderwal.net/random/entrysel.php?blog=1635. Zuletzt abgerufen am 19.8.2007.

Verzeichnis von Webadressen

[1] Flickr, http://www.flickr.com. Zuletzt abgerufen am 25.8.2007.

[2] Del.icio.us, http://del.icio.us. Zuletzt abgerufen am 23.8.2007.

[3] RealTravel, http://www.realtravel.com. Zuletzt abgerufen am 27.8.2007.

Semantic Web vs. Web 2.0

Armin Stein

1 Einleitung

Das als Internet bekannte Netzwerk wurde in seiner Entstehung lediglich als Ansammlung von Elementen gesehen, auf die in erster Linie von Menschen zugegriffen wird [3]. Anfangs handelte es sich bei den Inhalten um Texte und Bilder. Heute stellt sich das Angebot als wesentlich vielfältiger dar – Musik, Videos oder Spiele können ebenso über das Internet bezogen werden wie Informationen über das Wetter, Flugtickets oder Bücher. Der Zugriff auf die Angebote erfolgt mittlerweile nicht mehr nur durch Menschen, sondern in einer Zeit, in der das Internet nahezu ubiquitär erscheint, auch durch Computer und Anwendungen selbst. Während der Mensch noch die Möglichkeit hat, durch seine Interpretationsfähigkeit die Zusammenhänge und mögliche Verwendung der Inhalte zu verstehen, braucht ein Computer Regeln und Normen, welche diese Inhalte beschreiben und sie in einen semantischen Kontext führen. Das *Semantic Web* soll durch Ontologien die Bedeutung der Inhalte des World Wide Webs (WWW) hinterlegen, damit diese von Maschinen besser auswertbar und interpretierbar sind. Das Ziel dieser Ausarbeitung ist es, mögliche Verknüpfungen zwischen dem Semantic Web und dem Web 2.0 aufzuzeigen. Hierzu wird nach einer Begriffsabgrenzung (Kapitel 2) das Semantic Web in Bezug auf die Dimensionen *Technik*, *Ökonomie* und *Soziologie* vorgestellt. Im Anschluss daran werden einige Integrations- und Kombinationsmöglichkeiten der beiden Strukturen aufgezeigt (Kapitel 4), bevor ein Ausblick auf weitere Entwicklungsmöglichkeiten die Ausarbeitung schließt (Kapitel 5).

2 Begriffsabgrenzung

Semantic Web

Der ursprünglich von TIM BERNERS-LEE eingeführte Begriff *Semantic Web* [4] wird vom World Wide Web Consortium (W3C) als ein Framework definiert, das es ermöglicht, Daten – hauptsächlich, aber nicht notwendigerweise ausschließlich im Internet [7] – über Applikations-, Unternehmens- und Gemeinschaftsgrenzen hinweg zu verteilen und wiederzuverwenden [19]. Entgegen der ehemals üblichen und bis heute gebräuchlichen harten Verlinkung von Inhalten durch HTML (HyperText Markup Language), die durch Menschen erstellt werden kann und interpretiert werden muss, soll das Semantic Web durch maschinenlesbare Metainformationen die Möglichkeit schaffen, Inhalte so zu beschreiben, dass auch Maschinen sie interpretieren können – weshalb das daraus entstehende Netzwerk auch als *Web of Meaning* [11] bezeichnet wird. Über die Richtigkeit des Begriffs *Semantic Web* selbst wird derzeit diskutiert, da die Bezeichnung *Semantisches* Netz als nicht vollständig zutreffend angesehen wird [5]; als treffender wird der Begriff *Semiotic* Web erachtet, da die Semantik mit ihrem Fokus auf Sinn und Bedeu-

tung [8][1] lediglich ein Unterbereich der Semiotik ist, die das, was das Semantic Webs darstellen soll, durch ihre drei Elemente – Syntax, Semantik und Pragmatik – besser beschreibt [14]. Nachdem der von TIM BERNERS-LEE geprägte Begriff jedoch zumindest in der Informatik allgemeine Anerkennung gefunden hat, wird er in dieser Arbeit auch weiterhin im Sinne des W3C verwendet. Das sich entwickelnde Netzwerk beinhaltet demnach also Daten, die nicht nur hart syntaktisch verbunden sind, sondern über Metainformationen – gespeichert in Ontologien – miteinander in Beziehung stehen und somit zu weiterverwertbaren Informationen werden.

Web 2.0

Im Gegensatz zum Semantic Web gibt es für den Begriff *Web 2.0* (oder *Social Web*, vgl. bspw. [18]) keine einheitliche Definition wie im Falle des Semantic Web. Hervorgegangen ist der Begriff vielmehr aus einer Diskussion zwischen O'REILLY und MEDIALIVE INTERNATIONAL darüber, wie sich Unternehmen, die den Zusammenbruch des Aktienmarktes im Jahr 2000 überstanden, von jenen unterscheiden, die dabei untergegangen sind [13]. Hieraus resultierte eine Gegenüberstellung von Elementen der *post-crash*-Zeit, die als Merkmale des *Web 2.0* betitelt wurden und Elementen der *pre-crash*-Zeit, welche nachträglich *Web 1.0* genannt wurde. Dies lässt erkennen, dass die eigentliche Idee des Web 2.0 im Gegensatz zum Semantic Web nicht konkret auf einem Framework oder einer Technik beruht, sondern vielmehr ein Konglomerat von technischen Neuerungen und Ideen darstellt. Im Mittelpunkt dieser Innovationen steht beispielsweise der Ansatz des *User-Generated Content*. War es früher Spezialisten vorbehalten, Inhalte ins WWW einzustellen, ist es aufgrund der Vereinfachung der Anwendbarkeit der Technologien heute auch Amateuren möglich, sich im Netz zu präsentieren. Alles, was dafür benötigt wird, ist ein Internetzugang, dessen Verbreitung in den letzten Jahren stark zugenommen hat,[2] sowie ein Webbrowser. Dies ermöglicht die Verwendung bspw. von Blogs, elektronische Tagebücher, welche mittlerweile ohne Kenntnisse von Installationsmethoden kostenlos angelegt und verwaltet werden können.[3] Auch für die Verwaltung von Bildern steht kostenlos Speicherplatz zur Verfügung. Der Dienst flickr.com bietet bspw. einen kostenlosen Bilderupload von bis zu 100 MB pro Monat an – 2001 kosteten 6 MB Speicherplatz noch ca. 4 € pro Monat.[4] Über diese Bilder kann dann mittels Kommentarfunktionen diskutiert werden. Aber nicht nur im Amateurbereich findet dieser Diskurs statt. Auch die Presse verwendet von Nutzern erzeugte Informationen. Videos von medienwirksamen Ereignissen, die z. B. mit dem Mobiltelefon aufgenommen wurden, werden für Nachrichtensendungen aufbereitet und gesendet.[5] Somit kann das *Mitmach-Web*[6] als Grundlage und Motor einer benutzergetriebenen Erweiterung der Daten- und Informationsbasis des Internets gesehen werden.

---

[1] Nachzulesen auch unter http://www.ac-nancy-metz.fr/enseign/philo/textesph/frege_begriff_und_gegenstand.pdf, Abrufdatum: 2007-08-30.

[2] Im Jahr 2002 verfügten 43 % aller privaten deutschen Haushalte über einen Internetanschluss [15]; 2006 waren es bereits 61,4 % [16].

[3] Vgl. bspw. http://www.blogger.de, Abrufdatum: 2007-09-02.

[4] Vgl. bspw. http://web.archive.org/web/20011007070959/www.strato.de/visitenkarte/index.html, Abrufdatum: 2007-09-02.

[5] Vgl. bspw. I-Report beim Sender CNN: http://www.cnn.com/exchange/ireports/topics/forms/breaking.news.html, Abrufdatum: 2007-09-02.

[6] http://www.heute.de/ZDFheute/inhalt/25/0,3672,5252025,00.html, Abrufdatum: 2007-08-30.

3 Grundlagen des Semantic Webs

3.1 Technische Aspekte: Sprachen

Grundlagen – RDF

Mit XML (eXtensible Markup Language) existiert eine Sprache, die es ermöglicht, Elemente des Internets mit selbst definierten Tags zu beschreiben, welche durch Skripte und Programme extrahiert werden können. Trotzdem bleibt Systemen die Bedeutung des Inhalts verschlossen. Erst durch eine Sprache wie RDF (Resource Description Framework), die auf XML beruht und Tripel nutzt, die vergleichbar mit einer Satzkonstruktion – bestehend aus Subjekt, Prädikat und Objekt – sind, können Informationen sinnvoll nutzbar gemacht werden. Diese Statements haben die Form $A(O,V)$, was bedeutet, dass ein Objekt O ein Attribut A mit dem Wert (Value) V besitzt. Somit lassen sich einfache Beziehungen der Art „*http://www.uni-muenster.de*" „*repräsentiert*" „*Westfälische Wilhelms-Universität*" erzeugen. Eine alternative Darstellung ist $[O]–A\rightarrow[V]$, was als gerichtete und bewertete Kante zwischen zwei Knoten eines Baumes zu verstehen ist und dem Graphen-Charakter von RDF gerecht wird. Da innerhalb von RDF mit URI (Uniform Resource Identifier) gearbeitet wird, sind zumindest im Internet alle Objekte eindeutig adressierbar; es ist theoretisch aber auch möglich, andere URI-Schemata zu verwenden, sodass auch Elemente außerhalb des Internets beschrieben werden können [7]. Das folgende Beispiel versieht die Seite http://www.w3.org/RDF/ mit einem *title*, einem *publisher* und einem *contributor*. Es handelt sich dabei um eine Auswahl des *Dublin Core* (http://dublincore.org/), eine Übereinkunft auf 15 beschreibende Kernelemente einer Internetseite:

```
<?xml version = "1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
       xmlns:dc="http://purl.org/dc/elements/1.1/">
 <rdf:Description rdf:about="http://www.w3.org/RDF/">
  <dc:title>Resource Description Framework (RDF)</dc:title>
 </rdf:Description>
 <rdf:Description rdf:about="http://www.w3.org/RDF/">
  <dc:publisher>World Wide Web Consortium</dc:publisher>
 </rdf:Description>
 <rdf:Description rdf:about="http://www.w3.org/RDF/">
  <dc:contributor>Eric Miller</dc:contributor>
 </rdf:Description >
</rdf:RDF>
```

Obwohl durch RDF einige Merkmale der Seite beschrieben werden, gibt es noch keine Information über die Eigenschaften selbst oder deren Beziehungen untereinander. Aus diesem Grund wurde RDF Schema als semantische Erweiterung für RDF entwickelt, mit deren Hilfe Klassen und Eigenschaften durch ein erweitertes Vokabular beschrieben werden können. Dies wird durch das folgende Beispiel dargestellt:

```
...
<rdfs:Classrdf:ID="Person">
 <rdfs:comment>Personenklasse</rdfs:comment>
 <rdfs:subClassOfrdf:resource="http://www.w3.org/1999/02/22
       -rdf-syntax-ns#Resource"/>
</rdfs:Class>
<rdfs:Classrdf:ID="Student">
<rdfs:comment>Studentenklasse</rdfs:comment>
 <rdfs:subClassOfrdf:resource="#Person"/>
</rdfs:Class>
<rdfs:Classrdf:ID="Dozent">
...
```

Hier werden zwei Subklassen (Student, Dozent) der Klasse Person erzeugt, die wiederum eine Subklasse von Ressource ist. Somit besteht also die Möglichkeit, *Klassen* und *Unterklassen* sowie *Literale* (also Strings und Integers) zu bilden und diese untereinander zu verknüpfen. Diese Beziehung wird durch ein *property* zwischen einem Subjekt und einem Objekt erzeugt. Der folgende Code verbindet einen Kurs mit einem Dozenten:

```
<rdf:Property rdf:ID="dozent">
 <rdfs:comment>Dozent eines Kurses</rdfs:comment>
 <rdfs:domainrdf:resource="#Kurs"/>
 <rdfs:rangerdf:resource="#Dozent"/>
</rdf:Property>
```

Die hier mit RDF-S definierten Beziehungen sind ohne weiteren Kontext jedoch nicht zu verstehen, da über den Gegenstand der Objekte keine Information vorliegt. Diese soll über Ontologien geschaffen werden.

Ontologien – OWL

Wie bereits erwähnt, ist das Internet ein Medium, das zur Kommunikation zwischen mindestens zwei Partizipanten dient, wobei sicherzustellen ist, dass sich beide Seiten über das genutzte Vokabular einig sind. Wenn dies der Fall ist, gibt es aber noch keine Garantie, dass die Semantik der Daten, die sie zur Information macht, sowohl beim Sender als auch beim Empfänger die gleiche ist. Problematisch sind diese Schwierigkeiten vor allem bei der Kommunikation zwischen Maschinen: Aus was für einer Domäne die Variablenbezeichnung „name“ stammt, für die z. B. ein Service einen Wert erwartet, ist für einen anderen Dienst, der die Information liefern soll, ohne gewisse Übereinkünfte nicht erkennbar. Aus der Philosophie stammt zu diesem Problem der Begriff *Ontologie*, der sich mit dem Wesen des *Seins* befasst. Erst wenn Klarheit über das *Sein* eines Objektes herrscht, kann gesichert von ihm und über es geredet werden. Im Bereich der Informatik ist er derzeit noch nicht klar beschrieben. Eine der am häufigsten zitierten Definitionen stammt von GRUBER: „An ontology is an explicit specification of a conceptualisation“ [10], wobei es sich bei der Konzeptualisierung um Beziehungen von Objekten, Konzepten und anderen Gegenständen innerhalb einer gemeinsamen Domäne handelt. Sie stellt somit eine vereinfachte abstrakte Sicht auf alle Dinge und deren Beziehungen dar, die aus einem bestimmten Beweggrund abgebildet werden sollen. Als Sprache steht OWL (Web Ontology Language) zur Verfügung, welche die Konstrukte URI und RDF/RDF-S benutzt, um systemübergreifende Verteilung, Skalierung und Kompatibilität zu Webstandards sowie Erweiterbarkeit zu garantieren. Es ergänzt das vorhandene Vokabular von RDF-S beispielsweise um Kardinalitäten, Aufzählungsklassen und disjunkte Klassen. Somit kann eine OWL-Ontologie durch eine formale OWL-Semantik logische Konsequenzen erzeugen, die nicht in der Ontologie selbst vorhanden sind. Also kann durch Inferenz unter den Voraussetzungen „Peter ist Mensch“, „Menschen haben Beine“ geschlossen werden, dass Peter Beine hat. Hierdurch wird auch eine semantische Verknüpfung der Inhalte des WWW ermöglicht.

Kritische Betrachtung

Problematisch an den oben vorgestellten Technologien ist der Anwendungsbezug. Zur weit verbreiteten Nutzung muss sich auf einen gemeinsamen Nenner in Bezug auf die Anwendung bspw. der RDF-Elemente *Eigenschaften*, *Klassen* oder *Beziehung* geeinigt werden. Dies setzt einen sehr hohen Grad an und Bereitschaft zur Übereinkunft der Anwender voraus. In unterschiedlichen Anwendungsdomänen werden Inhalte verschieden beschrieben. VOSSEN und

HAGEMANN vergleichen die Problematik der Verbindung von Ontologien zur gemeinsamen Abdeckung von Domänen mit der Integration von Datenschemata unterschiedlicher Unternehmen [18]. Während das eine Unternehmen den Vornamen und Mittelnamen seiner Mitarbeiter als ein Datenfeld speichert, werden diese unter Umständen von einem weiteren Unternehmen in einem Feld gespeichert. Dennoch führt dies zu Überlegungen in Bezug auf Geschäftsmodelle des Semantic Webs, da die Möglichkeit, Daten zu beschreiben und auch im Web semantisch navigierbar zu machen, für unterschiedliche Unternehmen interessant sein kann.

3.2 Ökonomische Aspekte: Monetarisierung

TIM BERNERS-LEE verglich bereits 2001 die Möglichkeiten der Datenbankintegration mit der Integration von mit RDF beschriebenen Unternehmensdaten [2]. Die Integrationsproblematik der Datenbanken soll aufgrund gemeinsamer Beschreibungen der Vergangenheit angehören. Es wäre lediglich eine Metabeschreibung der in den unterschiedlichen Unternehmen vorhandenen Daten notwendig. Darüber hinaus sind – idealtypisch gedacht – alle Möglichkeiten der Kombination von im Web zur Verfügung gestellten Diensten möglich, um *Mash-Ups*, also komplexe Dienste zu erstellen. Bei einer einheitlichen Beschreibung der Schnittstellen von Web Services und der Registrierung bei den entsprechenden UDDI-(Universal Description, Discovery and Integration-)Servern wäre das Just-in-Time-Verbinden ohne menschliche Interaktion (Orchestrierung) möglich. Während einzelne Services für sich lediglich kleine Bereiche einer Aufgabe bearbeiten, können die Funktionalitäten komplexer Transaktionen, die als ein Dienst erscheinen, verkauft werden. HousingMaps[7] ist ein solcher Service, der Orte in den USA, an denen Immobilien angeboten werden, auf die Karten des Dienstes http://maps.google.com überträgt. Obwohl die Benutzung derzeit noch umsonst ist, wäre eine Verwertung als kostenpflichtiger Dienst denkbar. Aber auch Einzeldienste, die von einem Unternehmen zur Verfügung gestellt werden, können – mit den entsprechenden Informationen bezüglich des Preises versehen – gegen Bezahlung als Teil eines Mash-Ups genutzt werden. Infrage käme für einen solchen Dienst bspw. die Abfrage über Daten von Patenten beim Patentamt. Wiederum ist dieser Ansatz aufgrund der Notwendigkeit einer gemeinsamen Beschreibung und dem verbundenen Aufwand fraglich. Weiterhin besteht das Problem der Vorhersehbarkeit der Kostenentwicklung seitens des Nutzers. Wenn keine Kenntnis über die verwendeten Dienste besteht, kann ex ante keine Aussage bezüglich des zu erwarteten Preises getroffen werden.

3.3 Soziale Aspekte: Kommunikation

Die Technologien des Semantic Webs, wie sie sich derzeit präsentieren, zielen nicht darauf ab, die verbale Kommunikation zwischen Menschen zu unterstützen. Da die Metainformationen, die über im Web gespeicherte Elemente und Daten vorliegen, zur maschinenlesbaren Beschreibung derselben gedacht sind, wird der Bereich Mensch-zu-Mensch-Kommunikation eher ausgeblendet. Dies widerspricht jedoch nicht einem Potenzial, das mit der Existenz von Sprachen wie RDF-S oder OWL einhergeht. Der *Dublin Core* bietet die Möglichkeit, Metadaten über Personen abzulegen. Wenn diese in eine entsprechende Beziehung gebracht und mittels Inferenz ausgewertet werden, kann das Finden eines für eine spezielle Situation passenden Kontakts möglich sein. Sollten sich die Beschreibungselemente soweit erweitern lassen, dass allgemei-

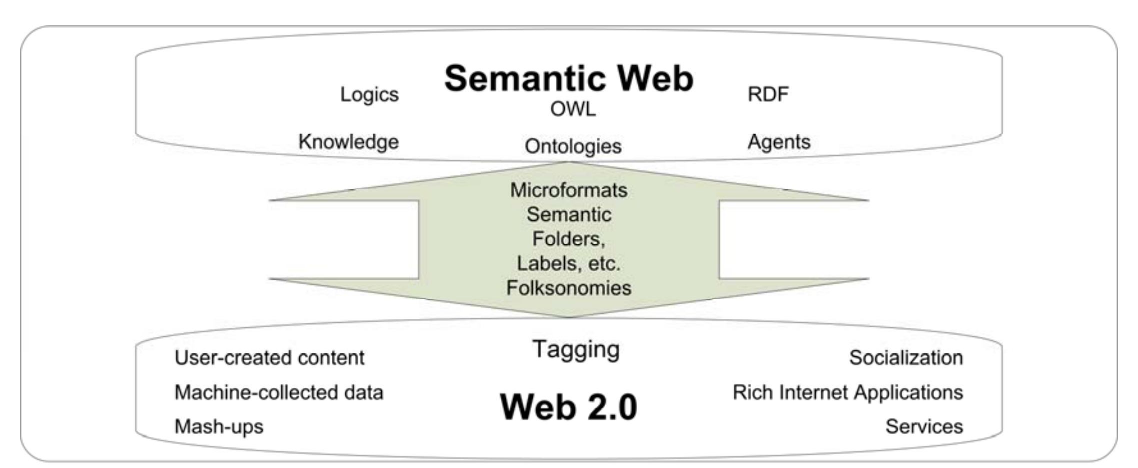
---

[7] http://www.housingmaps.com. Abrufdatum: 2007-08-31.

ner Konsens über die Bedeutung herrscht, wären – unter der Voraussetzung, dass im eigenen System Informationen über die Identität des Benutzers vorliegen – Suchen der Art „Welcher Zahnarzt in meiner Nähe kann Implantate einsetzen? “ möglich. Über die eigenen Informationen, die bspw. auf Clientseite abgelegt werden, kann – in Verbindung mit einem Dienst wie bspw. http://maps.live.com – der eigene Standort sowie der eines Zahnarztes identifiziert werden. Sollten Berufsbezeichnungen semantisch hinterlegt sein, wäre auch das Auffinden eines für diese Aufgabe *richtigen* Arztes unproblematisch.[8] Hier ist aber wiederum die Frage zu klären, aufgrund welcher Konventionen die Informationen für welche Elemente vergeben werden.

4 Semantic Web *vs.* oder *und* Web 2.0

VOSSEN UND HAGEMANN sehen das *Tagging*, d. h. das mit Metainformationen Versehen von Daten, in Verbindung mit Ontologien als die Schnittmenge oder zumindest Schnittstelle zwischen dem Semantic Web und dem Web 2.0 [18]. Abbildung 1 stellt diesen Bezug dar. Diese durch Darstellungen der vorhergehenden Abschnitte nachvollziehbare Aussage wird relativiert durch die Tatsache, dass auf der einen Ebene des Modells – auf der des Semantic Webs – strukturierte Metadaten vorliegen, die sich an Konventionen wie bspw. dem Dublin Core orientieren müssen, wenn Ontologien nicht mit viel Aufwand miteinander verbunden werden sollen. Auf der anderen Ebene – der Web 2.0-Ebene – liegen Unmengen an unstrukturierten Metainformationen durch Tags vor, die Benutzer an Elemente des WWW annotiert haben.



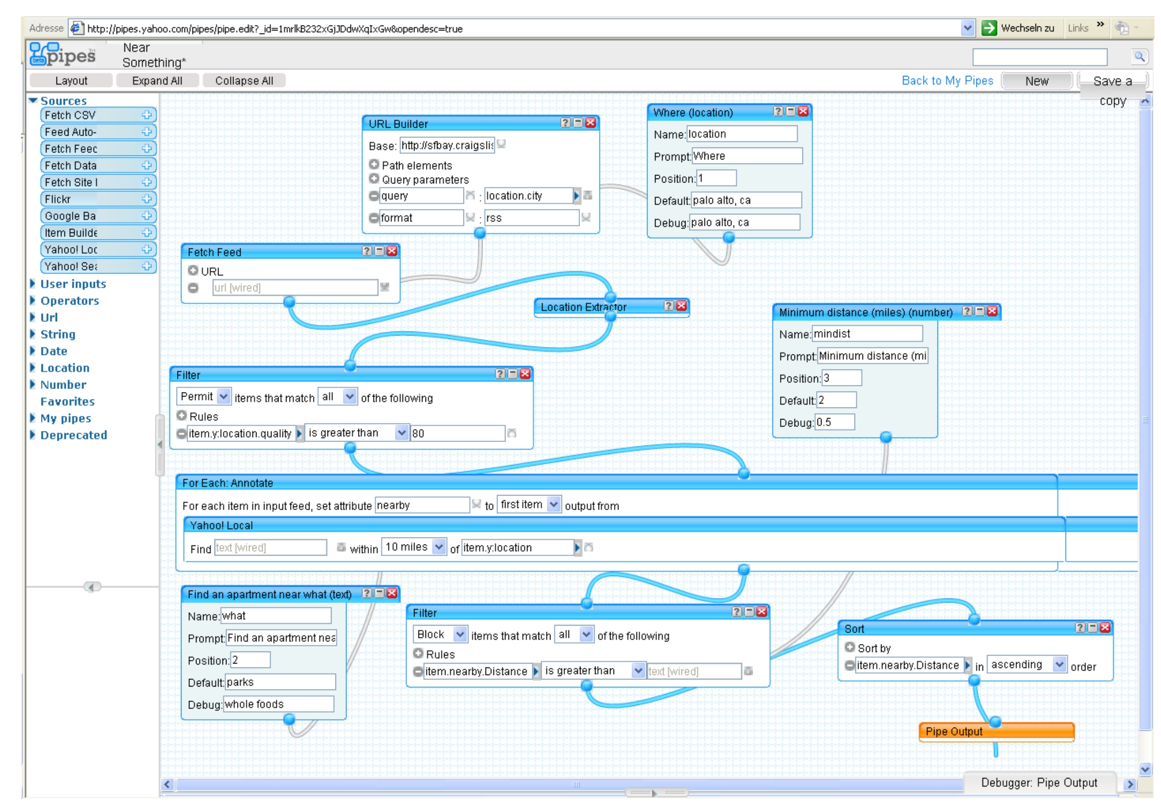
Quelle: [18], S. 336.

Abbildung 1: Zusammenhänge zwischen dem Web 2.0 und dem Semantic Web

Es gibt kaum Vorschriften, die festlegen, wie bestimmte Elemente getaggt werden sollen. Bei Flickr werden Tags z. B. frei in eine Zeile eingegeben und einem Bild zugeordnet. Welche Semantik aber dahinter steht, bleibt verschlossen. Eine Suche nach Bildern mit dem Tag „Dame“ kategorisiert nicht, ob es sich bei dem Bild um eine weibliche Person, Notre Dame in Paris oder ein Bild eines Dame-Spiels handelt. Trotzdem werden ca. 200.000 Treffer geliefert (Abrufdatum: 2007-09-05). Eine semantische Anreicherung der Daten um Domänen wäre hier sinnvoll und könnte es der Programmlogik der internen Suchmaschine erleichtern, die richtigen Ergebnisse

[8] Dieses Szenario ist angelehnt an [17] und die Ideen des FOAF-Projekts – http://www.foaf-project.org. Abrufdatum: 2007-08-30.

zu liefern. Die Clusterfunktion von Flickr geht in diese Richtung.[9] Eine weitere Alternative wäre das clientseitige Hinterlegen von persönlichen Informationen wie Interessensbereiche – wenn der Benutzer in seinem Profil angibt, dass er sich für Architekturfotografie interessiert, könnten ihm Frauenbilder gänzlich vorenthalten oder aber erst sehr spät in der Suchliste präsentiert werden. Eine ähnliche Entwicklung sehen ANKOLEKAR ET AL. [1], die in ihrem Positionspaper http://pipes.yahoo.com/ vorstellen, eine erste Kombination von Möglichkeiten des Semantic Webs – der Beschreibung von Inhalten, die aus dem Internet gelesen werden – mit der einfachen Bedienung des Web 2.0. Per *Drag&Drop* können elementare Funktionen wie *Date Input*, *Regex* oder *Fetch CSV* zu komplexen *Mash-Ups* kombiniert werden (vgl. Abbildung 2).[10]



Quelle: http://pipes.yahoo.com/

Abbildung 2: Beispiel eines Mash-Ups

Weiterhin existieren Suchmaschinen, die sich eine semantische Suche zunutzen machen, wie bspw. http://search.creativecommons.org, eine Metasuchmaschine über Google, Yahoo, Flickr und weitere. Sie liefert ausschließlich Ergebnisse, die frei verfügbar und nutzbar sind. Eine weitere, sich derzeit in der α-Testphase befindliche semantische Suchmaschine ist http://www.freebase.com. Hier werden Sucherergebnisse um semantische Informationen angereichtert, was bspw. bei der Suche nach *O'Reilly Media* als Ergebnis nicht nur Informationen über das Unternehmen, sondern auch das Faktum, dass es sich um ein solches handelt, liefert. Eine weitere Anreicherung des Web 2.0 durch Semantik kann durch die Strukturierung bereits vorhandener Daten durch Metadaten gesehen werden. Am Beispiel http://www.musicovery.com lässt sich erkennen, dass die Hinterlegung von Gemütszuständen und Musikrichtungen für Mu-

---

9 Vgl. http://www.flickr.com/photos/tags/BEZ/clusters/, wobei *BEZ* für einen zu clusternden Begriff steht.

10 Der entstandene Service kann unter http://pipes.yahoo.com/pipes/1mrlkB232xGjJDdwXqIxGw benutzt werden.

sikdateien eine Navigation durch den Suchraum bietet, der einen erstaunlich hohen Anteil an für den Moment *passender* Musik liefert. Musikdienste wie http://www.last.fm geben dem Benutzer die Möglichkeit, die gehörte Musik direkt über einen Web-Player zu bewerten und lassen diese Information in die Entscheidung über die folgenden Stücke einfließen. Über die eigenen gespeicherten Musikvorlieben können Teilnehmer gesucht werden, die einen ähnlichen Geschmack haben, sodass die Tracklisten gegenseitig ausgetauscht und erweitert werden können. Zusätzlich zu dem Musikangebot liefert last.fm mittels Google AdSense (https://www.google.com/adsense) Informationen über den Interpreten sowie Links zu Onlineshops, die den Künstler betreffende Tickets oder CDs verkaufen.

In Anbetracht der aufgezeigten Möglichkeiten müsste die Überschrift der Arbeit also „Semantic Web *und* Web 2.0“ lauten, da die Integration von Methoden des Semantic Webs in die vorhandenen Funktionalitäten des Web 2.0 zu neuen, ursprünglich undenkbaren – und teilweise auch aufgrund der überschaubaren Datenmenge damals vielleicht unnötigen – Möglichkeiten führt.

5 Ausblick

Während das Web 2.0 derzeit eher die soziale Komponente des Webs darstellt, indem es vielen Benutzer die (häufig unstrukturierte) aktive Teilnahme am Netz ermöglicht, befasst sich das Semantic Web mit der Bedeutungsebene und der Verbindung von Daten in strukturierter Weise. Diese zwei kontroversen Elemente verfügen in ihrer Kombination aber über das große Potenzial, eine weitere Generation des Internets zu gestalten, die teilweise bereits mit *Web 3.0* bezeichnet wird ([6], [9], [12]). Könnte die durch die Menge der Benutzer anhaltende Datenflut strukturiert mit Metadaten, die bspw. durch *Folksonomies* – d. h. gemeinschaftliches Taggen von Daten – zur Verfügung gestellt werden, angereichert werden, wäre eine einfachere, durch Systeme unterstützte Navigation durch den Raum des WWW möglich. Hierbei ist es jedoch als wichtig zu erachten, dass das Erstellen der Metadaten für den Benutzer ähnlich einfach und intuitiv sein muss, wie die Benutzung des Web 2.0 mit seinen aktuellen Möglichkeiten. Dies ist aber, wie VOSSEN UND HAGEMANN darstellen, keine triviale Aufgabe ([18], S. 310 ff.). Aus diesem Grund scheinen die Möglichkeiten, die eine Konzentration auf die Entwicklung der Integration der beiden Netze hervorbringen kann, den Aufwand lohnenswert zu machen. Erste Bestrebungen gibt es z. B. in der Form von *Microformats*, mit deren Hilfe RDF-ähnliche Informationen direkt in HTML-Seiten integriert werden können (http://www.microformats.org). Derzeit existieren Vorlagen für z. B. Lebensläufe (*hResume*), elektronische Visitenkarten (*hCard*) oder Kalenderdaten (*hCalendar*). Ausgehend von diesen können in Zukunft weitere Bereiche erschlossen werden.

Literaturverzeichnis

[1] A. Ankolekar, M. Krötsch, T. Tran, and D. Vrandecic, The two cultures – mashing up web 2.0 and the semantic web, in Proceedings of the World Wide Web Conference, 2007.

[2] T. Berners-Lee, Business model for the semantic web. http://www.w3.org/DesignIssues/Business.html, 2001, Abrufdatum: 2007-09-01.

[3] T. Berners-Lee, R. Fielding, and H. Frystyk, Hypertext transfer protocol – http/1.0. http://www.ietf.org/rfc/rfc1945.txt, 1996, Abrufdatum: 2007-08-31.

[4] T. Berners-Lee, J. Hendlerand, and O. Lassila, The semantic web. URL: http://www.sciam.com/print_version.cfm? articleID=00048144-10D2-1C70-84A9809EC588EF21, 2001, Abrufdatum: 2007-09-01.
[5] A. Blumauer and T. Pellegrini, Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen, Springer, Berlin Heidelberg New York, 2006, ch. 1, pp. 9-24.
[6] M. V. Copeland, What's next for the internet. http://money.cnn.com/magazines/business2/business2_archive/2007/07/01/1%00117068/ index.htm?postversion=2007070305, July 2007, Abrufdatum: 2007-08-30.
[7] D. Fensel, O. Lassila, F. van Harmelen, I. Horrocks, and J. Hendler, The semantic web and its languages, IEEE Intelligent Systems, 15 (2000), pp. 67-73.
[8] G. Frege, Sinn und Bedeutung, Zeitschrift für Philosophie und philosophische Kritik, 100 (1892), pp. 25-50.
[9] B. Getting, Basic definitions: Web 1.0, web 2.0, web 3.0. http://www.practicalecommerce.com/articles/464/Basic-Definitions-Web-10%-Web-20-Web-30/, July 2007, Abrufdatum: 2007-08-30.
[10] T. R. Gruber, A translation approach to portable ontology specification, Tech. Report KSL 92-71, Stanford University, Stanford, California 94305, 1992.
[11] P. Marsh, Web of meaning, IEE Computing & Control Engineering, April/May (2004), pp. 10-15.
[12] P. Midwinter, Is google a semantic search engine? http://www.readwriteweb.com/archives/is_google_a_semantic_search_engine%.php, March 2007, Abrufdatum: 2007-08-30.
[13] T. O'Reilly, What is web 2.0. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web%-20.html, 2005, Abrufdatum: 2007-09-01.
[14] J. F. Sowa, Ontology, metadata, and semiotics. http://www.jfsowa.com/ontology/ontometa.htm, 2000, Abrufdatum: 2007-08-30.
[15] Statistisches Bundesamt, Informationstechnologie in unternehmen und haushalten 2005. http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/C%ontent/Publikationen/Fachveroeffentlichungen/Informationsgesellschaft/pressebr%oschuere__ikt2005,property=file.pdf, Wiesbaden, 2006, Abrufdatum: 2007-09-02.
[16] Statistisches Bundesamt, Private haushalte in der Informationsgesellschaft – Nutzung von informations- und Kommunikationstechnologien (ikt). Wirtschaftsrechnungen 15, 2007.
[17] K. Tochtermann and H. Maurer, Semantic Web – Geschichte und Ausblick einer Vision, 2007, ch. 0, pp. 1-6.
[18] G. Vossen and S. Hagemann, Unleashing Web 2.0 – From Concepts to Creativity, Elsevier, Inc., 2007.
[19] World Wide Web Consortium, Semantic web. http://www.w3.org/2001/sw/, 2001, Abrufdatum: 2007-08-30.

User-Generated Advertising

Burkhard Weiß

1 User-Generated Advertising – Hype oder Zukunftsmodell?

1.1 Motivation – Der „Nutzer“ und UGA im Mittelpunkt

Das Time Magazin kürte am Ende des Jahres 2006 die Person des Jahres 2006. Diese war jedoch nicht wie üblich eine bestimmte Person, die in dem Jahr vielfach medial präsent war, sondern „You“. Damit wollte das Times Magazin auf die immer größer werdende Rolle des Nutzers in der öffentlichen Meinungsbildung aufmerksam machen. Auch das Magazin Advertising Age kürte Anfang 2007 die Agentur des Jahres 2006 – diese war jedoch ebenfalls nicht wie sonst eine der bekannten international tätigen Werbeagenturen, sondern in Anlehnung an die Time Magazin Wahl „der Konsument“, der zunehmend in der Öffentlichkeit auch als Werbeproduzent agiert.

Mit den einhergehenden starken Veränderungen des Internets in den letzten Jahren von einem reinen Zugriffsmedium auf ein interaktives Medium unter dem Stichwort „Web 2.0“, gerät somit der Nutzer immer stärker in den Fokus, da er nicht mehr nur konsumiert, sondern zunehmend Meinungen verbreitet und Inhalte selbst produziert. Dieser Wandel hin zu einer Verbraucherdemokratie führt dazu, dass auch Unternehmen sich zunehmend mit den Konsequenzen auseinandersetzen müssen und ihre Kommunikation und insbesondere ihre Werbeaktivitäten neu gestalten müssen. Der Konsument, der täglich einer enormen Werbeflut ausgesetzt ist und allmählich werberesistent geworden ist [vgl. o. V. (AMA) 2006 sowie Seith 2007, denenzufolge 95 % aller täglichen Werbebotschaften nicht wahrgenommen werden], muss auf neuen Wegen von den Unternehmen angesprochen und gebunden werden, um deren langfristiges Bestehen und deren Erfolg am Markt zu sichern.

Ein erster Lösungsansatz bzw. Ausweg aus dieser Krise wird seit 2006 zunehmend unter dem Begriff „User-Generated Advertising“ (UGA) in der Werbebranche forciert. In diesem Dokument soll daher eine kritische Untersuchung dieser neuen Entwicklung stattfinden und daraus folgend sollen Implikationen für die Zukunft aufgezeigt werden.

1.2 Gegenstandsbereich des UGA

Mit der Veränderung des Internets zu einem partizipativen Medium und dem Rollenwandel des Konsumenten hin zu einem aktiven Nutzer (der selbst User-Generated Content produziert) müssen auch Werbestrategien neu ausgerichtet werden. Während früher Werbung vor dem Hintergrund einer spezifischen Zielgruppe mit einer bestimmten Werbebotschaft erstellt wurde, muss Werbung zukünftig den Konsumenten bzw. Nutzer aktiv einbinden und dahin führen, dass der Nutzer zum Werbetreibenden für das Unternehmen wird [vgl. Vossen, Hagemann 2007, S. 276], da Unternehmen ihre ursprüngliche Macht und Rolle als Markengestalter mit dem Wandel der Nutzer in der Verbraucherdemokratie verloren haben.

Diese Idee – den Nutzern die Gestaltung der Werbung im Einzelnen oder als Gesamtes zuzusprechen – findet ihren Niederschlag u. a. im derzeitigen anglo-amerikanischen Hype-Begriff des „User-Generated Advertising" (UGA).[1] User-Generated Advertising soll hier wie folgt charakterisiert werden [vgl. auch abweichend hierzu Wunsch-Vincent, Vickery 2007, S. 4, 8]:

- eine kreatives Werk bzw. eine kreative Leistung (Werbung),
- die außerhalb einer professioneller Tätigkeit erstellt wird,
- mit dem Ziel für ein bestimmtes Produkt, eine Dienstleistung, Organisation, Person oder Marke zu werben.

Die Formen von User-Generated Advertising werden dabei meist klassisch unterteilt in textuelle Werbung, Fotos und Grafiken, Musik und Audio, sowie Video bzw. Film [Wunsch-Vincent, Vicekery 2007, S. 16 f.]. Kanäle die UGA nutzt sind neben dem Internet als ursprüngliches Medium, insbesondere Printmedien, das Radio und Fernsehen.

Ziel des UGA ist es aus Sicht der Werbewirtschaft, dass sich Kunden durch Mitmachwerbung mit dem Beworbenem intensiv beschäftigen, Ideen generieren, ihre Begeisterung für das Beworbene ausbauen und im Idealfall die selbst erstellte Werbung oder auch Werbebotschaften im Sinne des Beworbenen selbst verbreiten [vgl. hierzu bspw. Seith 2007, die hierbei von einem „viralen Effekt" spricht].

1.3 Zielstellung – UGA auf dem Prüfstand

Seitdem die, durch UGA forcierte neue Leitidee der Werbewirtschaft – Konsumenten zu Werbetreibenden zu machen – im Jahr 2006 in den USA vermehrt Anwendung gefunden hat und im Positiven wie im Negativen in die Schlagzeilen geraten ist, stellt sich die Frage welche Schlüsse aus den bisherigen Entwicklungen zu ziehen sind, welche Chancen und Risiken UGA nachweislich bietet und welche Erfolgsfaktoren die Anwendung von UGA-Konzepten unterstützen. Aus einer derartigen Analyse sind sodann aktuelle und zukünftige Entwicklungsrichtungen abzuleiten und Perspektiven für den weiteren Einsatz von UGA und nicht zuletzt die Werbewirtschaft aufzuzeigen. Dies soll in den beiden folgenden Kapiteln erfolgen, um abschließend einen Ausblick in Form von zusammenfassenden Thesen zu geben.

2 Bisherige Entwicklungen zum UGA im Web 2.0 und deren Bewertung

2.1 Entwicklungstreiber des UGA – Vom Konsumenten zum Prosumer

Um die aktuellen und bisherigen Entwicklungen um UGA herum nachvollziehen zu können, bedarf es zunächst eines Rückblicks auf die wesentlichen Treiber, die das Aufkommen von

---

[1] Neben User-Generated Advertising ist auch eine Vielzahl an Synonymen, insbesondere Consumer Generated Advertising, User-Created Advertising, Participatory Advertising, Crowd Generated Advertising, User-Generated Branding in diesem Kontext gebräuchlich. Verwandte Begriffe umfassen darüber hinaus insbesondere Citizen Marketing (mit dem Fokus auf Bürger und „Markenevangelisten" [McConnell, Huba 2007, S. 4 und McConnell, Huba 2005]), Social Advertising (mit dem Fokus Werbung in sozialen Netzwerken) und Brand Democratization (mit dem Fokus der Demokratisierung von Marken).

UGA begünstigt haben. Treiber der UGA-Entwicklung lassen sich dabei in vier Kategorien unterteilen: technologische Treiber (bspw. Verbreitung von Breitbandzugängen und neue Webtechnologien), soziale Treiber (bspw. Umgang mit persönlichen Daten und Bedürfnis, das eigene Leben auch online auszuleben), ökonomische Treiber (bspw. das gestiegene Interesse an den Möglichkeiten der Kommerzialisierung des Internets) und institutionelle Treiber (bspw. die Entstehung neuer, flexiblerer Lizenzschemata, wie die Creative Common Licence) [vgl. auch Wunsch-Vincent, Vickery 2007, S. 13 f., die diese Treiber im Rahmen von User-Created Content identifizieren]. Diese Treiber werden im Folgenden in Abbildung 1 kurz charakterisiert.

Technologische Treiber:

- erhöhte Speicher- und Prozessorleistung
- erhöhte Qualität von audiovisueller Technologie bei Verbrauchern
- Entstehung einfacher Technologien und Software zur Kreation, Verbreitung und gemeinsamen Teilung von Inhalten
- erhöhte Breitbandzugänge zum Internet

Soziale Treiber:

- Entstehung einer neuen Zielgruppe von jungen, internetaffinen, engagierten u. technisch versierten Nutzern
- Verbreitung von Inhalten, Empfehlungen und Bewertungen sowie freiwillige Veröffentlichung persönlicher Informationen im Internet
- zunehmender Wunsch, eigene Individualität auszuleben; Wunsch nach Interaktivität
- Entstehung von Communities und kollaborativen Projekten
- Politik, Bildung, Bedürfnis nach sozialem Engagement

Ökonomische Treiber:

- günstigere Software zur Erstellung von medialen Inhalten
- Finanzierung von und Investments in User-Created Content-Seiten
- günstiger Internetzugang und Breitbandanschluss
- Interesse an User-Created Content seitens kommerzieller Akteure
- erhöhte Ausgabebudgets für Werbung im Internet

Institutionelle Treiber:

- Entwicklung von flexiblen Lizenzen für den Zugang und die Weiterverarbeitung v. Werken
- Entstehung v. Endnutzerlizenzvereinbarungen, die Nutzern Urheberrechte zugestehen

Abbildung 1: Treiber der UGA-Entwicklung

In Folge der Auswirkung dieser Treiber (im Rahmen des „Web 2.0“) entwickelte sich der Konsument im Internet zusätzlich zum Produzenten von Inhalten im Internet. Diese neue, erweiterte Rolle des Nutzers wird festgehalten im Begriff des Prosumer[2] [vgl. Tapscott, Williams 2006, S. 125] – ein Kunstwort, das die eigentlich konträren Begriffe des Produzenten und Konsumenten zu vereinen sucht.

---

2 Neben dem Begriff „Prosumer“ ist auch der Begriff „ProdUser“ als Kunstwort aus Produzent und Nutzer von Inhalten verbreitet.

Mit der Evolution des Konsumenten zum Prosumer entsteht auch aufseiten der Werbetreibenden das Bedürfnis diese neue Zielgruppe hinsichtlich der eigenen Marke adäquat zu adressieren, um so Aufmerksamkeit und eventuell auch weiteres Interesse beim Nutzer für die eigene Marke zu erzeugen. Eine adäquate Zielgruppenansprache im Rahmen des User-Generated Advertising gelingt jedoch nur dann, wenn zuvor die Motive des neu definierten Nutzers identifiziert werden können, die diesen bewegen und motivieren. Grundsätzlich lassen sich dazu sowohl intrinsische als auch extrinsische Motive feststellen [vgl. Weber 2006, S. 29 f. sowie Zouain 2006, S. 35]. Intrinsisch motiviert sind demnach Nutzer, wenn sie von der Herausforderung oder auch der „Spaß an der Sache" getrieben werden und daraufhin eigenes Engagement entwickeln. Extrinsisch motiviert ist der Prosumer hingegen bspw. dann, wenn er getrieben wird von der Möglichkeit, Reputation bzw. gesellschaftliche Anerkennung zu gewinnen, enthusiastisch für die Marke ist, sich mit der Gruppe der Kreativen identifiziert, aus Altruismus heraus handelt oder auch nach finanziellen Vorteilen strebt. Dieser so charakterisierte Prosumer ist der Kernakteur des UGA-Ansatzes, der im Folgenden aus Sicht der Werbetreibenden und Unternehmen und deren Erfahrungen mit dem Prosumer im Rahmen von UGA-Kampagnen näher untersucht werden soll.

2.2 Unternehmen auf dem Weg zum User-Generated Advertising

Unternehmen haben auf die Veränderung des Konsumenten zum Prosumer reagiert und testen seither unterschiedliche Arten des User-Generated Advertisings. Ziel ist es dabei einerseits, den Kunden entgegenzukommen und diese aktiv in die Markengestaltung einzubinden und andererseits, den größtmöglichen Nutzen (bspw. in Form von zusätzlicher kreativer Werbung, Mund-zu-Mund-Propaganda und medialer Aufmerksamkeit) aus dieser Entwicklung bzw. Kommunikationsform mit den Kunden zu ziehen.

Die Ursprünge des User-Generated Advertising werden oft den USA zugeschrieben [vgl. Karig 2007, S. 14]. Ein früher Prototyp ist demzufolge die Mozilla Foundation gewesen, die auch für die Werbung ihres populären Firefox-Browsers ihrer Open Source-Philosophie treu blieb und öffentlich ihre erste große Fernsehkampagne ausschrieb. Zu dieser konnte jeder – ob Fan, Profi oder Laie – beitragen und der gewinnende Fernsehspot wurde sodann veröffentlicht.

Seitdem setzen viele meist amerikanische Unternehmen verstärkt auf Wettbewerbe, um von ihren Kunden gestaltete Werbefilme für ihre Marke zu erhalten, die dann bspw. im Fernsehen, auf Messeauftritten oder zu großen medialen Ereignissen (wie etwa dem Super Bowl in den USA) ausgestrahlt werden. Dabei wird den Kunden jedoch teilweise nicht nur die Erstellung der Werbung ermöglicht, sondern es wird etwa auf Konzepte des Crowdsourcings [vgl. hierzu Whitla 2007] zurückgegriffen, um Nutzer auch an der finalen Entscheidung über die zu sendende Werbung teilhaben zu lassen. Im Idealfall verbreitet der Nutzer darüber hinaus bspw. über positive Mund-zu-Mund-Propaganda die so gewählte Markenbotschaft in Form von Linkempfehlungen zu den betreffenden Werbefilmen auf Videoplattformen wie YouTube.

Je nach Vertrauen der Unternehmen in ihre Marke und Kunden werden derartige Wettbewerbe teilweise sehr restriktiv und teilweise sehr offen ausgeschrieben. So schrieb Chevrolet bspw. den Wettbewerb „The Apprentice" aus, bei dem es seinen Kunden online verschiedene Werbesequenzen für das Modell Chevrolet Tahoe nach dem Baukastenprinzip vorgab und die Kunden bat, diese zusammenzufügen und mit Texten und Musik zu hinterlegen. Umweltaktivisten, die

gegen den dort zu bewerbenden Geländewagen waren, nutzen jedoch die restriktive Plattform, um ihrerseits zynische Filme zu schneiden. Chevrolet veröffentlichte jedoch nur ausgewählte Videos (nicht die 16 % der negativen Videos von bspw. den Umweltaktivisten [Simms 2007]) und versuchte, die Kontrolle über die Markenkampagne zu behalten. Die Umweltaktivisten verbreiteten jedoch ihre geschnittenen Videos auf anderen Internetplattformen und erlangten durch ihre Parodien eine viel größere Öffentlichkeit, als die Kampagne ohne sie gehabt hätte. Dies ist sicherlich das bekannteste Negativbeispiel einer UGA-Kampagne.

Andere Kampagnen, wie etwa die Converse Gallery, waren offener im Umgang mit Ihrer Marke und überließen den Kunden die volle Kontrolle über die Inhalte des Werbefilms und sogar die Entscheidung, ob die zu bewerbenden Converse-Schuhe überhaupt in den Videofilmen zu sehen waren. Die Resonanz seitens der begeisterten Converse-Kunden war groß und wurde zum Erfolgsbeispiel einer UGA-Kampagne.

Auch deutsche Unternehmen versuchen zunehmend UGA-Kampagnen für ihre Zwecke einzusetzen. So schrieb die BMW-Tochter Mini bspw. einen Onlinewettbewerb für die Einführung neuer Modelle aus, da man mit einem um 75 % gekürzten Werbebudget auskommen musste. Die Ergebnisse überraschten die Marketingabteilung, denn viele Einreichungen kamen von passionierten Mini-Fans, die insbesondere auch durch ihre Kreativität bei der Umsetzung der 30-Sekunden-Werbefilme überzeugen konnten.

Die wohl derzeit meist erwähnte UGA-Kampagne war der Wettbewerb von Doritos, der für den Super Bowl – das mediale Sportereignis des Jahres 2007 in den USA – einen Wettbewerb für die Öffentlichkeit ausschrieb. Der Werbefilm des Siegers, der von fünf Personen innerhalb von vier Tagen für 12,79 $ gedreht wurde, wurde sodann auf dem teuersten Sendeplatz des Jahres während des Sportevents ausgestrahlt.

UGA-Kampagnen werden jedoch nicht immer nur von Firmen initiiert, sondern gehen teilweise auch von deren Kunden direkt oder indirekt aus. So ist hier insbesondere das Beispiel zweier Komiker zu nennen, die aus Cola Light-Flaschen und Mentos in vielen belegten Experimenten Hunderte von Geysiren simulierten. Während die Coca Cola-Firma diese Verschwendung ihres Getränks als imageschädigend ansah, war Mentos erfreut über die unerwartete und vielfach verteilte und Aufmerksamkeit erregende Videowerbung und sponserte das Team mittlerweile mehrfach.

Negative Auswirkungen hatte insbesondere eine UGA-Kampagne, die einen Terroristen in einem VW Polo zeigte. Dieser ist zu sehen, wie er ein Autobombenattentat durchführen will und dabei nur selbst im Innenraum des Fahrzeugs durch die Explosion getötet wird, während das äußere des Fahrzeugs nicht beschädigt wird und somit die Menschen herum unverletzt bleiben. Aufgrund der Professionalität des Videos bezichtigte man Volkswagen einer geschmacklosen Werbekampagne, obwohl Volkswagen den Spot weder in Auftrag gegeben noch selbst hergestellt hatte. Es stellte sich heraus, dass der Spot zwar von einer professionellen Werbeagentur gedreht wurde, jedoch ohne die Zustimmung von Volkswagen.

Für Aufsehen sorgte auch ein Mann in der Schweiz, der vorgab, für Marken wie Gucci und Armani zu modeln und professionell gestaltete Werbeanzeigen von sich und den Marken in Zeitschriften schaltete. Die Werbeanzeigen schaltete er jeweils kurz vor Anzeigenschluss, sodass die Zeitungen keine Zeit hatten, die Echtheit des Kontakts zu überprüfen. Dabei gab er zur Ab-

rechnung der Kosten als Auftraggeber jeweils die richtigen Anschriften der Modemarken an, die jedoch zuvor nichts mit den Anzeigen zu tun hatten.

Die letzte, hier zu behandelnde, von Nutzern initiierte UGA-Kampagne stellt kein weiteres Beispiel aus der Wirtschaft dar, sondern beschäftigt sich mit der Politik – genauer der amerikanischen Präsidentschaftswahl 2008. Anfang des Jahres 2007 geriet ein politisch motiviertes Video an die Öffentlichkeit, das Hillary Clinton als Diktatorin in Anlehnung an George Orwell's Buch „1984“ zusammen mit Auszügen aus ihrer Präsidentschaftsrede portraitiert. Am Ende des 74-sekündigen Videos wird ein Text eingeblendet, dass Anfang 2008 die Wahlen zur Präsidentschaft sind und 2008 nicht so sein wird wie 1984. Das Video endet mit dem Verweis zur Homepage des Gegenkandidaten Barack Obama. Zwei Wochen wurde gerätselt, wer der Produzent dieses Videos war, der Hillary Clinton in ein schlechtes Licht stellte und die Marketingstrategen des Gegenkandidaten Obama gerieten unter Verdacht. Es stellte sich heraus, dass ein überzeugter Obama-Fan und Mitarbeiter einer Werbeagentur der Demokraten in seiner Freizeit das Video erstellt hatte.

Aus all diesen Beispielen ist ersichtlich, dass sich UGA als Werbeformat erst in seinem Anfangsstadium – der Experimentierphase – befindet. Eines zeichnet sich jedoch schon ab: während Unternehmen sich eingestehen müssen, die mediale Kontrolle über das Markenimage zu verlieren, ist der Prosumer bereits dabei, das Markenimage in vielen Bereichen aktiv mitzugestalten – ein Wandel, der aus aktueller Sicht unumkehrbar scheint.

2.3 Vor- und Nachteile und Chancen und Risiken des UGA

User-Generated Advertising ist ein Werbeformat, das – wie die Beispiele gezeigt haben – viele Chancen und Risiken mit sich bringt und etliche Vor-, aber auch Nachteile hat. Bevor Unternehmen von sich aus UGA-Kampagnen initiieren, sollten daher kritisch alle Argumente, die für und wider derartige Kampagnen sprechen, sorgfältig gegeneinander abgewogen werden. Auf Basis einer Literaturanalyse[3] wurde daher im Rahmen der vorliegenden Arbeit eine Argumentenbilanz erarbeitet, die als Entscheidungsgrundlage für die Durchführung einer User-Generated Advertising-Kampagne dienen kann (vgl. Abbildung 2). Diese Argumente gelten jedoch nur eingeschränkt bzw. in Abhängigkeit von der konkreten Ausgestaltung der UGA-Kampagne (restriktiv vs. nicht-restriktiv, Nutzer produzieren nur Werbung oder verteilen diese auch ...).

---

[3] Der Literaturanalyse wurden diverse Quellen aus Zeitschriften, Journals und anderen Veröffentlichungen im Internet zugrunde gelegt [zu den Quellen vgl. bspw. Arora, Jethi 2006; Karig 2007; Moskowitz 2006a; Moskowitz 2006b; Mullich 2007; o. V. (AMA) 2006; Pfanner 2006; Story 2007; von Puttkamer 2007; Wegert 2006; Whitla 2007; Wunsch-Vincent, Vickery 2007 sowie Zouain 2006].

Vorteile:	**Nachteile:**
■ günstigere Werbeproduktion, da von Nutzern erstellte Werbung ■ Kunde beschäftigt sich mit Marke. ■ Kunde identifiziert sich mit Marke. ■ Kunde „lebt" das Markengefühl in der Werbung aus. ■ Kundenengagement durch Kundeneinbindung in die Werbeproduktion und eventuell -auswahl ■ Ausnutzung der Authentizität des Prosumers ■ höhere Glaubhaftigkeit bei Zielgruppe, weil Werbung von Nutzern produziert wurde ■ Möglichkeit der Messung der Passion der Marke bei Kunden ■ Möglichkeit der Messung des Werbeerfolgs bei Kunden über Downloads, Views etc. ■ hohe Qualität der Kundenkontakte, da hauptsächlich markeninteressierte Nutzer	■ Kontrollverlust der Werbetreibenden über die Inhalte der Werbebotschaft bzw. sogar der ganzen Kampagne ■ hohe Kosten durch Kampagnenplanung, -begleitung, insbesondere Koordination der Einreichungen und Selektion der Einreichungen, sowie bei Veröffentlichung in traditionellen kostenpflichtigen Medien ■ Verlust eines einheitlichen Markenimages ■ Rechte an Werbeinhalten (insb. Nutzungs-/Verwertungs- und Schutzrechte)
Chancen:	**Risiken:**
■ Abschöpfung von externer Kreativität ■ verstärkte Öffentlichkeitswirkung ■ Erschließung neuer Zielgruppen ■ Erreichung junger, internetaffiner, werberesistenter Nutzer ■ Möglichkeit, Werbetalente zu entdecken ■ Multiplikatoreffekte durch Verteilung der Werbung durch werbebegeisterte Nutzer ■ Möglichkeit, neue Kundenkontakte zu generieren ■ Möglichkeit, Kundenmeinungen zu sammeln und Kommunikation über Marke zu forcieren	■ Aufdeckung von Mängeln (insbesondere an Produkten und Dienstleistungen) ■ negative Beiträge von Markengegnern ■ Werbeüberflutung durch Generierung vieler Beiträge ■ mangelnde Qualität der Beiträge (technisch, aber auch inhaltlich) ■ Verlust einer markenkonformen Werbebotschaft ■ Haftung für Werbeinhalte ■ Erreichung falscher Zielgruppen ■ Verwässerung der Marke

Abbildung 2: Argumentenbilanz zum User-Generated Advertising

Hervorzuheben sind insbesondere die folgenden Vorteile einer UGA-Kampagne. Eine derartige Kampagne geht einher mit einer hohen Qualität an Kontakten und nutzt nicht nur den Multiplikatoreffekt bei Nutzern, sondern auch die Möglichkeit der Weiterempfehlung: Empfänger bewerten demzufolge virale Videos nicht als unaufgeforderte Werbung, wenn sie von Freunden zugesandt werden, sondern als Unterhaltung [Kiley 2005], die eher nebenläufig auch für eine bestimmte Marke etc. wirbt. Auch können so insbesondere ansonsten werberesistente Nutzer erreicht werden [vgl. Whitla 2007, S. 6 f.]. Des Weiteren kann die Begeisterung der Nutzer für die Marke genutzt werden, um neue Zielgruppen zu erschließen, Medienaufmerksamkeit zu generieren und allgemein Konsumenteninteresse zu erzeugen [vgl. Moskowitz 2006b]. Vor allem die Authentizität der Nutzer [vgl. dazu Moskowitz 2006b], die in der Werbung reflektiert wird, erzeugt eine hohe Akzeptanz der Werbung bei der Zielgruppe [Tomczak, Schögel, Sulser 2006, S. 76] bzw. begründet ein verstärktes Vertrauen in die jeweilige Marke [vgl. die Studienerkenntnisse der American Marketing Association – o. V. (AMA) 2006]. Kundenbindung und Loyalität hingegen werden durch die aktive Einbindung der Nutzer in die Werbung gefördert

[Wegert 2006]. Meist wird durch derartige Kampagnen die Förderung des Communitygedanken und der Netzwerkbildung unter den Nutzern noch gestärkt, da sie gegenseitig die Möglichkeit haben, sich miteinander auszutauschen. Außerdem werden häufig Freunde und Bekannte bei der Produktion und Beurteilung der Werbung mit einbezogen und durch die aktive Beschäftigung mit der Marke und den Markenwerten entsteht eine hohe Akzeptanz oder auch Identifikation mit der Kampagne oder sogar auch der Marke [vgl. Tomczak, Schögel, Sulser 2006, S. 76]. Auch besteht die Möglichkeit, neue Kundenkontakte zu akquirieren [Tomczak, Schögel, Sulser 2006, S. 76], wenn Nutzer sich für derartige Wettbewerbe registrieren müssen. Über die Möglichkeit, auf Formularen E-Mail-Adressen von Freunden und Bekannten einzutragen, um diesen originelle oder innovative UGA-Werbeanzeigen zuzuschicken (bzw. Links zu diesen), ergibt sich eine weitere Möglichkeit, neue Kontakte zu schaffen. Auf diese Arten erzeugte Kontakte sind dabei meist wertvoller als große, aufgekaufte Kontaktdatenbanken (deren Kontakte nicht bereits vorselektiert wurden bzw. auf die Marke positiv reagieren).

Zu den bedeutenden Nachteilen einer derartigen Kampagne gehört oft auch der einhergehende Kontrollverlust über die offizielle Werbebotschaft bzw. Verwässerung dieser [vgl. Zouain 2006, S. 34 und von Puttkamer 2007] und das Risiko der Aufdeckung von Mängeln an bspw. Produkten oder Dienstleistungen [Wunsch-Vincent, Vickery 2007, S. 34 f.]. So wurde der Fall eines Herstellers von Sicherheitsschlössern in den Medien bekannt, bei dem ein Nutzer ein Video ins Internet stellte, in dem er demonstrierte, wie diese Schlösser des Herstellers mit einfachsten Mitteln sekundenschnell geöffnet werden konnten. Subversion, Adbusting und Markensatire können somit insbesondere für große Marken, wie Shell, Nike und McDonalds, die von der Öffentlichkeit kritisch betrachtet werden, zu einem nicht zu unterschätzenden Risiko für UGA-Kampagnen sein [vgl. Karig 2007, S. 16 sowie Wegert 2005b]. Vielfach werden auch die Kosten und der Aufwand von UGA-Aktivitäten unterschätzt. So lassen sich zwar Kosten in der Werbeproduktion einsparen, jedoch entstehen vielfach Kosten durch die nötige Koordination der meist zahlreichen Einreichungen und Begleitung der UGA-Kampagne [vgl. Story 2007]. So muss nicht nur technologisch eine Plattform für die Einreichungen bereitgestellt werden, sondern ein UGA-Wettbewerb muss auch beworben werden und Einreichungen müssen bearbeitet sowie alle Beiträge zu einer Kampagne zusammengeführt werden [Marchese 2007]. Eine vorherige Bestimmung des zu erwartenden Werbeerfolgs ist darüber hinaus schwierig, wenn überhaupt möglich [vgl. Mullich 2007, S. 7]. Auch das Problem der Rechte an den Videos, inbesondere auch der verwendeten Musik stellt UGA-Kampagnen vor nicht unerhebliche Probleme [Wunsch-Vincent, Vickery 2007, S. 43-51]. Insbesondere die Erstellung und Veröffentlichung illegaler, anstoßender oder nicht gewünschter Inhalte in Videos können bei UGA-Inhalten rechtliche Konsequenzen nach sich ziehen [Wunsch-Vincent, Vickery 2007, S. 54].

Die Selektion der „richtigen“ (von Nutzern akzeptierten, von Marke gewünschten, rechtlich unproblematischen etc.) Werbung unter allen nutzergenerierten Inhalten ist eines der größten Probleme [vgl. Hempel 2007 sowie Whitla 2007, S. 9], denen man sich auf Plattformen für User-Generated Content (UCA) insbesondere mittels drei verschiedener Verfahren stellt [Wunsch-Vincent, Vickery 2007, S. 53 f.]:

- Moderation vor Veröffentlichung: Experten treffen die finale Entscheidung über die Veröffentlichung von UGA-Inhalten.
- Moderation nach Veröffentlichung: Nach der Veröffentlichung können Moderatoren ggf. die UGA-Inhalte abändern oder entfernen.

- Moderation während Veröffentlichungsprozess: Sog. Peers – also Mitbewerber – übernehmen die Bewertung von UGA-Inhalten und entscheiden in Summe über das Ranking der UGA-Werbeeinreichungen.

Neuere Ansätze zur Verwaltung und Selektion in UGA-Kampagnen hingegen greifen auf die Konzepte der sozialen Filtration und gegenseitigen Bewertung der Nutzer untereinander (Reputationsmechanismen) und nicht nur ihrer gelieferten Inhalte zurück. Für Nutzer und letztendlich auch Werbetreibende (sofern sie ihre Werbeversprechen erfüllen können) wird darin eine große Chance zur Steuerung von UGA-Kampagnen gesehen, da dadurch eine möglichst faire Bewertung und Verbreitung dieser Meinungen gefördert wird.

Aktuelle Studienergebnisse der American Marketing Association [o. V. (AMA) 2006] bescheinigen UGA-Kampagnen jedoch überwiegend einen positiven Nutzen. So werden Firmen, die UGA einsetzen – im Gegensatz zu solchen, die nur auf professionelle gestaltete Werbung vertrauen – in einer Befragung von 1.098 Subjekten als kundenfreundlicher (68 %), kreativer (56 %) und innovativer (55 %) eingestuft, wobei jüngere Subjekte zwischen 18-24 Jahren Unternehmen, die UGA einsetzen, etwas kritischer bewerten als Subjekte zwischen 25-64 Jahren.

Letztendlich muss jedoch neben allen genannten Aspekten beachtet werden, dass der Erfolg dieses neuen Marketinginstruments teilweise noch von der Neuheit des Ansatzes profitiert [vgl. Moskowitz 2006a] und Zusammenhänge zwischen den Ausgestaltungen von UGA-Kampagnen und spezifischen Vor- und Nachteilen bzw. Chancen und Risiken noch wenig erforscht sind.

2.4 Erfolgsfaktoren und Erfolgsmessung des UGA

Obwohl UGA-Kampagnen eine Reihe von Vor- und Nachteilen besitzen, wie im vorangegangen Kapitel erläutert wurde, wurde bisher noch nicht weiter untersucht, was bestimmte UGA-Kampagnen erfolgreich bzw. erfolgreicher als andere macht, die eventuell sogar negative „Publicity" erzeugen.

Erfolgreiche Kampagnen hängen zunächst einmal von zwei Akteursgruppen maßgeblich ab: einerseits von den Teilnehmern an der UGA-Kampagne und andererseits von dem Werbenden bzw. der koordinierenden Organisation im Hintergrund der Kampagne.

So lässt sich feststellen, dass besonders solche UGA-Kampagnen erfolgreich sind, die Konsumenten mit einer Begeisterung für die Marke haben, sowie solche, die kreativ sind, die Zeit haben und über das technische Wissen verfügen, um eigene qualitative Beiträge liefern zu können [vgl. Bosman 2006, Moskowitz 2006a, Tomczak, Schögel, Sulser 2006, S. 76]. Dazu ist es selbstverständlich förderlich bzw. teilweise notwendig, dass die Marke bereits weitgehend bekannt ist, herausragend ist, Akzente setzt und mit den angebotenen Produkten und Dienstleistungen etc. die Zielgruppe überzeugen kann.

Auf der anderen Seite hängt der Kampagnenerfolg auch maßgeblich vom werbenden Unternehmen ab. So sind insbesondere solche Kampagnen erfolgreich, die sehr gut geplant bzw. vorbereitet wurden und professionell koordiniert werden [vgl. Roskos, Feldmann 2007], um aus der Vielzahl der Beiträge einen Nutzen zu ziehen. So wurde bspw. für den sehr erfolgreichen UGA-Wettbewerb der BMW-Tochter Mini mehrere Monate im Voraus eine crossmediale Kampagne mit Anzeigen in Lifestyle-Magazinen wie Fit for Fun, Max, Playboy sowie Webspecials

bei GQ und ProSieben geschaltet, um die Werte und Attribute der Marke Mini bekannter zu machen [Tomczak, Schögel, Sulser 2006, S. 74]. Jedoch nicht nur die Kampagnenplanung und -betreuung sind von großer Wichtigkeit für den Werbeerfolg, sondern auch die Inhalte und Rahmenbedingungen des Wettbewerbs selbst spielen eine große Rolle. So sind insbesondere solche Kampagnen erfolgreich, die den Nutzer in seiner Kreativität wenig oder gar nicht einschränken oder nur wenige Vorgaben geben und evtl. sogar dazu beitragen, eine markenaffine Community um die Kampagne herum aufzubauen. Positive Beispiele sind hier die zuvor erwähnte Converse Gallery, die über 400.000 Nutzer pro Monat anzog und 12 % Umsatzgewinn im Folgequartal bescherte, oder aber auch Nike mit der UGA-Kampagne „The Chain" [vgl. Simms 2007]. Nike erstellte bspw. eine Webseite und bat fußballbegeisterte Nutzer aus aller Welt, Videos von sich und ihren Künsten im Umgang mit dem Fußball einzusenden, wobei die einzige Vorgabe war, dass am Anfang jedes Videos ein Fußball von links kommen musste, der am Ende nach rechts aus dem Bild geschossen wurde. Durch die Aneinanderreihung der Videos ergab sich das „längste von Nutzern generierte Fußballvideo" der Welt, zu dem mehrere Hundert Videos aus über dreißig Ländern beitrugen. An dieser Kampagne spiegelt sich auch ein weiterer Erfolgsfaktor einer guten UGA-Kampagne wider: Erfolgreiche UGA-Kampagnen schaffen es nicht nur, alleine Werbung zu liefern, sondern auch den Teilnehmern (Prosumern) einen Dienst bzw. Mehrwert anzubieten, der als Anreiz dient, um mitzumachen. Bei Nike war dies demzufolge die Möglichkeit, seine Fußballkünste einem weiten Publikum unter Beweis zu stellen und so als Nutzer „soziale Anerkennung" zu erlangen. Ziel ist es daher, den Beitrag des Nutzers in den Mittelpunkt zu rücken [Wegert 2005a], sodass Nutzer und Werbetreibende gleichermaßen von der UGA-Kampagne profitieren. Die Nike (The Chain)-Kampagne ist ebenfalls ein erfolgreiches Beispiel dafür, wie dem Nutzer nicht nur gleichzeitig ein Dienst geboten wurde, sich der Öffentlichkeit zu präsentieren, sondern auch ein weiterer Anreiz geboten wurde, nämlich Teil einer einmaligen Aktion zu sein [Wegert 2005a] – der längsten Fußballsequenz der Welt.

Selbstverständlich trägt nicht zuletzt – neben der Herausforderung der Erstellung von guten Inhalten – auch ein gut gewählter Anreiz für die Nutzer dazu bei, qualitativ hochwertige Beiträge zu erhalten. So hat sich aus finanzieller Sicht ferner ein Preisgeld oder anderer materieller Anreiz in der Ausschreibung von Wettbewerben bewährt sowie die Möglichkeit, seine eigene Werbeidee professionell umzusetzen, oder bei einem exklusiven Markenevent (wie dem Super Bowl oder das Drehen eines echten Werbespots etc.) dabei sein zu können. Bei der BMW-Mini-Kampagne war bspw. entscheidend, ein BMW-Mini-Auto selbst als Gewinn anzubieten, da so eher echte Fans für den Wettbewerb begeistert werden konnten als professionelle Agenturen, die Geld vorziehen würden [vgl. Tomczak, Schögel, Sulser 2006, S. 76].

Neben den beiden Akteursgruppen – den Teilnehmern und den Werbetreibenden – spielt auch das Thema bzw. der Werbegegenstand (ob Marke, Produkt, Dienstleistung, Person, Organisation etc.) eine große Rolle, denn eine gute Kampagne schafft es, über eine kritische Masse an Einreichungen hinaus, die nötig ist um die Zielgruppe entsprechend breit anzusprechen und involvieren [vgl. Marchese 2007]. So liefern insbesondere solche UGA-Kampagnen die erfolgreichsten nutzergenerierten Werbebotschaften, deren Themen die Menschen begeistern; Beispiel hierfür sind etwa Politik und Sport [Mullich 2007, S. 7 sowie Wittrock 2007]. Als Beispiel kann hier wiederum das bereits genannte, politisch motivierte Video „Vote different" gegen Hillary Clinton als Präsidentschaftskandidatin angeführt werden, sowie ein UGA-Wettbewerb gegen Präsident Bush und seine Wiederwahl 2004, bei dem das Video „Child's Pay" von Charlie Fi-

sher gewann, dessen Spot jedoch dann nicht wie geplant während dem Super Bowl gezeigt werden durfte.

Obwohl die Erfolgsaussichten einer Kampagne vorab kaum abzusehen sind, bietet die Erfolgsmessung, im Gegensatz zu traditionellen Werbeformen (wie etwa Printanzeigen in Zeitungen oder Werbesequenzen im Fernsehen), deutlich vielfältigere Möglichkeiten. So können u. a. folgende Aspekte einer UGA-Kampagne gemessen werden:

- Zielgruppenerreichung: über die Art der Reaktionen von Nutzern und die Sammlung von Daten bei Registrierungsprozessen zum Wettbewerb über die Nutzer;
- Nutzerengagement: über die Zahl der Einreichungen und Qualität der Einreichungen;
- zeitliche Kampagnenauswirkung: über die mediale Präsenz bzw. Berichterstattung über die Kampagne und die Anzahl an Verlinkungen bzw. Referenzierungen in der Blogsphäre im Zeitverlauf;
- Verbreitung: über die Anzahl an Downloads, Anzahl an Besuchern einer Webseite, Anzahl an „Tell a Friend"-Funktionsnutzungen auf einer Webseite, die bei der Distribution der Werbebotschaften unter Freunden hilft;
- Interesse: über Anzahl an Teilnehmern sowie Anzahl an Downloads und Weiterempfehlungen, aber auch direktes Feedback an den ausrichtenden Werbetreibenden;
- positive Werbewirkung: über direkte Reaktionen per E-Mail, in der professionellen Berichterstattung oder über Einträge auf Blogs;
- Kaufwille: über bspw. die Verlinkung von Werbeanzeigen zu Produktseiten;
- Markenwahrnehmung: über Attribute, mit denen Prosumer die Marke in ihren Beiträgen bewusst oder unbewusst charakterisieren.

Zusammenfassend sind UGA-Kampagnen demnach nicht für jeden Werbegegenstand geeignet, sondern themenabhängig und ihr Erfolg hängt sowohl von der Marken-Community als auch von der werbenden Partei ab. Vorab lassen sich die Erfolgsaussichten nur teilweise abschätzen, aber die Erfolgsmessung ist umso vielfältiger möglich.

3 Aktuelle und zukünftige Entwicklungen des UGA

3.1 Traditionelle Werbeabteilungen und -agenturen im Umbruch

Mit der Verbreitung von UGA-Initiativen im Bereich der Werbung fürchten traditionelle Agenturen und Werbeberater, -designer, -planer und -strategen um ihre Zukunft. Traditionell konsultierten Organisationen diese, während die Arbeit der professionell arbeitenden Marketingabteilungen und Agenturen darin bestand, die Werbebotschaft in Abstimmung mit dem Klienten zu entwickeln, die eigentliche Werbekampagne und alle zugehörigen Medien zu entwerfen und zu verteilen. Die Rolle des Konsumenten bestand in der einseitigen Aufnahme der Werbung, der eventuellen Weiterempfehlung des Beworbenen an andere Konsumenten und schließlich in der Reaktion in Form von Investitionen (Käufen, Spenden etc.) in das jeweilig beworbene Objekt [vgl. Story 2007]. Mit dem Aufkommen des Prosumers verändert sich die traditionelle Rolle der

Werbeabteilungen und -unternehmen, indem sich die kreative Leistung und evtl. sogar Distribution der Werbung hin zum Prosumer verlagert. Mittlerweile nutzen nicht nur Unternehmen die Kreativität und den Arbeitswillen der Prosumer für die Werbung, sondern auch Werbeagenturen selbst direkt (wie bspw. die Agentur TBWA aus London auf ihrer Webseite unter der Initiative „The big what adventure" bekannt gab), indem sie die Nutzer um Feedback und Empfehlungen bitten, zu eigenen Beiträgen ermutigen und sogar einen finanziellen Anreiz für neue Ideen bieten [Whitla 2007, S. 7].

In Anlehnung an Marchese bieten sich traditionellen Werbeagenturen drei primäre Möglichkeiten, um den aktuellen Wandel hin zu UGA bewältigen zu können [vgl. Marchese 2007]:

- Das Anbieten oder Fördern von komplementären Dienstleistungen (wie etwa die Kampagnenplanung und/oder -betreuung),
- der Wandel der Agentur, sodass diese selbst als „Nutzer" an Wettbewerben teilnehmen kann und Einreichungen vornimmt; hierbei ist das Ziel, möglichst hoch dotierte Preisgelder zu gewinnen und gleichzeitig effizient und kostengünstig an Wettbewerben teilzunehmen;
- nachweisliche Verbesserung des eigenen Preis-Leistungs-Verhältnisses, um mit Kosten und Werbeerfolgen von UGA-Kampagnen mithalten zu können.

Insbesondere der erste Ansatz – als einziger Ansatz, der auf „neue" Marktlücken durch den Einsatz von UGA abzielt – verspricht zukünftig ein hohes Erfolgspotenzial, da aktuell beobachtet werden kann, wie am Markt neue und innovative Intermediäre entstehen, die sich insbesondere auf die Ausrichtung von UGA-Wettbewerben spezialisiert haben. Diese sollen im folgenden Kapitel näher untersucht werden.

3.2 Crowdsourcing-Plattformen – Koordinierende Werbeintermediäre

Während traditionelle Unternehmen der Werbebranche sich mit den neuen Möglichkeiten des UGA arrangieren müssen, entstehen aktuell neue Plattformen im Internet, deren Betreiber als Intermediäre zwischen Marken und Prosumer agieren.

Entsprechend dem Beispiel aus anderen Bereichen der Wirtschaft, wo Crowdsourcing (die Nutzung, Bündelung und Koordination kollektiver Intelligenz) bspw. in der Forschung und Entwicklung über die Internetplattform InnoCentive bereits erfolgreich eingesetzt wird,[4] bieten die neuen Intermediäre im Marketingbereich Unternehmen an, für sie UGA-Kampagnen umzusetzen, während sie den Kunden/Konsumenten/Prosumern/Nutzern anbieten, die Werbung selbst zu erstellen und zu verbreiten. Populäre Beispiele hierfür sind etwa die Unternehmen UGENMedia und ViTrue sowie Shycast, die Unternehmen flexible Wettbewerbsplattformen bieten, die diese an ihre Unternehmenswebseiten anbinden und auch nutzen können, um UGA-Wettbewerbe und -Kampagnen selbst effizient steuern und betreiben zu können. Weitere Beispiele stellen die Intermediäre Brickfish, CurrentTV, Revver, moviebakery (aus Deutschland), Shycast und zooppa dar, die selbst als Portal UGA-Inhalte verbreiten und ihren Nutzern u. a. für erstellte Werbespots eine finanzielle Entlohnung bzw. variable Vergütung in Abhängigkeit von der Popularität

---

[4] Diese neuen Crowdsourcing-Plattformen werden auch als Ideagoras bezeichnet [vgl. hierzu Tappscott, Williams 2006, S. 98].

der produzierten Videos anbieten. Teilweise bieten diese Intermediäre auch gleichzeitig die Technologien und Tools an, um auch komplexe Werbeinhalte wie Videos webbasiert erstellen zu können.

Das Geschäftsmodell dieser Intermediäre ist einfach und sehr effektiv: Unternehmen schreiben Wettbewerbe aus bzw. wenden sich an den UGA-Plattformintermediär, der dann die eigentlichen Crowdsourcing-Aktivitäten steuert. Dazu zählt insbesondere die Akquisition von Nutzern für das Portal und das Anbieten von Anreizen für diese, um sie für die Erstellung und Distribution von UGA-Inhalten zu gewinnen. Nutzer registrieren sich sodann und reichen eigene Beiträge ein. Zugleich können Nutzer über die Einreichungen anderer Nutzer abstimmen, wodurch nicht nur der Produktions-, sondern auch der Selektionsprozess an die Nutzer ausgegliedert wird. Je nach Popularität der Inhalte, erhalten die am besten bewerteten Einreicher fixe Preisgelder oder sogar in Abhängigkeit von der Anzahl an Downloads und der Verbreitung der Werbebeiträge variable Vergütungen.

3.3 Vom UGA zum Social Shopping

Die Weiterentwicklung dieses noch jungen Trends hin zu Crowdsourcing-Intermediären für die Werbung hat bereits einen neuen Niederschlag mit den Stichworten Social Advertising, Social Shopping und Social Commerce gefunden. Dabei geht es nicht mehr nur darum, die Produktion und anschließende Verbreitung von Werbung zu fördern, sondern insbesondere darum, durch Werbung innerhalb sozialer Netzwerke Kontakte zu einem Kauf zu führen.

Prominentes Beispiel hierfür ist das Social Shopping Portal Shoppero.com. Hier kann der Nutzer nicht nur die Werbung für ausgewählte Produkte in Form von so genannten Adgets[5] selbst erzeugen, sondern ist auch für die Verbreitung bspw. auf seinem eigenen Blog zuständig. Darüber hinaus werden die so erstellten Werbeanzeigen idealerweise mit Produkten des Herstellers verknüpft, die der Nutzer auf seinem Blog präsentiert, näher erläutert und empfiehlt. Ziel ist es, dass Leser bzw. Besucher derartiger privater Webseiten und Blogs überzeugt werden von authentischen und ehrlichen Empfehlungen, die sie im Anschluss zu einem Kauf des beworbenen Produkts verleiten. Das Konzept User-Generated Advertising wird also konsequent weiterentwickelt zum Social Advertising (Werbung innerhalb sozialer Netze) und im besten Falle ausgebaut zum Social Commerce (Verkauf von Produkten innerhalb sozialer Netze im Internet über persönliche Empfehlungen). Social Shopping-Portale bieten dabei den Nutzern die Technologie, um soziale Netzwerke aufzubauen und für Produkte in diesen aktiv zu werben. Der Anreiz für die Prosumer als Ersteller und Verteiler der Werbung wird bspw. dadurch gegeben, dass diese an den Werbeeinnahmen des Social Shopping-Portalbetreibers beteiligt werden. Bei Shoppero erhält der werbetreibende Nutzer bspw. 20 % der Werbeeinnahmen, wenn andere Nutzer seine private Webseite bzw. seinen Blog etc. lesen und 60 % der Werbeeinnahmen, sofern Nutzer auf seinem Blog auf das Adget klicken und so direkt zum Produktanbieter verlinkt werden. Aus traditioneller Sicht sind User-Generated Advertising und Social Shopping somit nur neue Ausprägungen des Prinzips „Kunden werben Kunden“ – mit dem Unterschied, dass sie auf die Kommunikation und Werbung im Internet sowie die Zielgruppe der internetaffinen Nutzer maßgeschneidert sind.

---

5 Adgets sind durch Nutzer erstellte Werbeboxen/Werbecontainer für Webseiten, die aus Bildern und Text bestehen und bspw. als Diashow innerhalb des Adgets abgespielt werden können.

4 Ein zusammenfassender Überblick und Ausblick

Zusammenfassend lässt sich festhalten, dass User-Generated Advertising nicht nur ein kurzweiliger Trend zu sein scheint, sondern ebenfalls eine neue Möglichkeit zur Betreibung von Werbung speziell im Internet für internetaffine Zielgruppen und unter Zuhilfenahme moderner Technologie und webbasierter Kommunikations- und Interaktionsmöglichkeiten. Mit jeder neuen UGA-Kampagne gewinnt dieser Trend immer mehr an Bedeutung und hat bereits jetzt schon die Zukunft der Werbewirtschaft, mit der Entstehung neuer Intermediäre, nachhaltig beeinflusst. Nicht nur die Werbung von Produkten, Dienstleistungen, Unternehmen und Marken in der Wirtschaft nutzt diese neue Werbeform, sondern auch in anderen Bereichen, wie der Politik und bei nicht-kommerziellen Projekten und Organisationen bietet User-Generated Advertising eine neue Option, Menschen für die verschiedensten Ziele, Visionen und Ideen zu erreichen. Zukünftig wird die Entwicklung jedoch nicht nur die Produktion und Distribution der Werbung einschließen, sondern im Rahmen von Social Shopping-Ansätzen wird es neue Möglichkeiten geben, den Prosumer auch über die Werbegestaltung hinaus einzubinden und über ihn neue Kanäle zum Verkauf von Produkten und Dienstleistungen zu erschließen. Der erfolgreiche Einsatz dieser neuen Form der Werbung hängt jedoch – im Gegensatz zu früher – nun hauptsächlich von den neuen produzierenden und vernetzten Konsumenten (Prosumer) ab [vgl. Zunke 2006], die zukünftig auch offiziell in der Werbung „das letzte Wort“ haben werden.

Literaturverzeichnis

Arora, S.; Jethi, M.: User Generated-Ads. 31.12.2006. http://www.onlinevideopunch.com/blog/index.php/user-generated-ads.htm, Abrufdatum: 3.9.2007.

Bosman, J.: User-generated content starts to take hold in advertising. 11.5.2006. http://www.iht.com/articles/2006/05/11/business/ads.php, Abrufdatum: 3.9.2007.

Hempel, J.: Crowdsourcing: milk the masses for inspiration, in: Business Week. 25.9.2006. http://www.businessweek.com/magazine/content/06_39/b4002422.htm, Abrufdatum: 3.9.2007.

Karig, F.: Jeder kann Werbung, in: Brandeins 9 (2007) 6, S. 14-16.

Kiley, D.: Advertising By, Of, And For The People. 25.7.2005. http://www.businessweek.com/magazine/content/05_30/b3944097.htm, Abrufdatum: 3.9.2007.

Marchese, J.: The Myth Of 'User'-Generated Advertising. 27.3.2007. http://publications.mediapost.com/index.cfm?fuseaction=Articles.showArticle&art_aid=57691, Abrufdatum: 3.9.2007.

McConnell, B.; Huba, J.: Welcoming citizen marketers into your world. 02.2005. http://customerevangelists.typepad.com/blog/2005/02/citizen_markert.html, Abrufdatum: 3.9.2007.

McConnell, B.; Huba, J.: Citizen Marketers: When People Are the Message. New York 2007.

Moskowitz, R.: Are Consumer-Generated Ads Here to Stay? 10.5.2006a. http://www.imediaconnection.com/content/9521.asp, Abrufdatum: 2.9.2007.

Moskowitz, R.: The Agency of the Future: Consumers. 26.4.2006b. http://www.imediaconnection.com/content/9265.asp, Abrufdatum: 2.9.2007.

Mullich, J.: Don't Monkey Around. User Generated Advertising Is Serious Business, in: Fuel (2007) 1. S. 6-7.

o. V. (AMA): Consumers Like Companies That Let Them Create Ads, But Young Adults Still Not Buying It. 1.12.2006. http://www.businessweek.com/the_thread/blogspotting/ archives/document/AMA %20release.doc, Abrufdatum: 3.9.2007.

Pfanner, E.: Leave It to the Professionals? Hey, Let Consumers Make Their Own Ads. 4.8.2006. http://www.nytimes.com/2006/08/04/business/media/04adco.html?ei=5088 &en=cb4c6604da68212a&ex=1312344000, Abrufdatum: 2.9.2007.

Roskos, M.; Feldmann, F.: Pro und contra zum Thema User generated advertising. 6.3.2007. http://www.visualblog.de/?p=513, Abrufdatum: 3.9.2007.

Seith, A.: Neuer Reklametrend. Werbespots zum Selberdrehen. 8.6.2007. http://www.spiegel.de/wirtschaft/0,1518,487205,00.html, Abrufdatum: 3.9.2007.

Simms, J.: And now a word from our customers…. 31.5.2007. http://www.redorbit.com/news/ technology/846978/and_now_a_word_from_our_customers/index.html, Abrufdatum: 3.9.2007.

Story, L.: The High Price of Creating Free Ads. 26.5.2007. http://www.nytimes.com/2007/05/26/ business/26content.html?ei=5088&en=f5244987dc59d9d0&ex=1337832000, Abrufdatum: 2.9.2007.

Tapscott, D.; Williams, A. D.: Wikinomics – How Mass Collaboration Changes Everything, New York 2006.

Tomczak, T.; Schögel, M.; Sulser, C.: Do-it-Yourself-Werbung für Mini, in: Persönlich – Die Zeitschrift für Unternehmensführung, Marketing und Kommunikation (2006) 9, S. 74-78.

von Puttkamer, R. K.: User generated advertising. 3.3.2007. http://www.werbeblogger.de/2007/ 03/03/user-generated-advertising/, Abrufdatum: 3.9.2007.

Vossen, G.; Hagemann, S.: Unleashing Web 2.0 – From Concepts to Creativity. Amsterdam 2007.

Weber, D.: Open Source Marketing. Master Thesis, Berlin School of Economics, Berlin 2006.

Wegert, S.: Is User-Generated Advertising Safe? 9.11.2006. http://www.clickz.com/3623884/print, Abrufdatum: 3.9.2007.

Wegert, S.: Media Revolution: Brand Democratization, Part 1. 4.8.2005a. http://www.clickz.com/3524906/print, Abrufdatum: 3.9.2007.

Wegert, S.: Media Revolution: Brand Democratization, Part 2. 11.8.2005b. http://www.clickz.com/ 3526556/print, Abrufdatum: 3.9.2007.

Whitla, P.: 'Crowdsourcing' and its Application in Marketing Activities, in: Proceedings of Business and Information 4 (2007) 1, Tokyo 13.7.2007.

Wittrock, P.: Anti-Clinton-Spot bringt Obama in Bedrängnis. 22.3.2007. http://www.spiegel.de/netzwelt/web/0,1518,473303,00.html, Abrufdatum: 2.9.2007.

Wunsch-Vincent, S.; Vickery, G.: Participative Web: User-Created Content. (OECD Report). 12.4.2007. http://www.oecd.org/dataoecd/57/14/38393115.pdf, Abrufdatum: 3.9.2007.

Zouain, R. R.: AS ARMAS DA RAPOSA. Como os novos produtores de conteúdo estão mudando a comunicação – e o que a publicidade tem a ver com isso. 10.11.2006. http://pack2go.files.wordpress.com/2007/03/tcc_roberta_screen.pdf, Abrufdatum: 3.9.2007.

Zunke, K.: Tyrann Kunde, in: ADZINE – Magazin für Online-Marketing. 24.11.2006. http://www.adzine.de/de/site/contentfcmsv44565bd054051a/artikel.xml, Abrufdatum: 3.9.2007.

Qualitätsaspekte im User-Generated Content

Jens Feldkamp

1 Einleitung

In diesem Beitrag wird der im Internet verfügbare, so genannte „User-Generated Content“ (UGC) in Bezug auf seine Qualität diskutiert. Aufgrund des begrenzten Raumes können hier nur ausgewählte Aspekte des Themas behandelt und selbst diese vielfach nur kurz angeschnitten werden; die vorliegende Darstellung kann daher keinen Anspruch auf Vollständigkeit erheben.

Die Kapitel 2 und 3 stellen zunächst die beiden Begriffe „User-Generated Content“ und „Qualität“ sowie deren Beziehung zueinander vor.

Kapitel 4 beschreibt anschließend einige ausgewählte Folgen, die insbesondere eine niedrige Qualität von UGC haben können, bevor im Kapitel 5 beispielhaft drei UGC-Plattformen behandelt werden. Hier werden insbesondere die Bemühungen der Plattformbetreiber zur Qualitätssicherung und deren Erfolg beleuchtet.

Kapitel 6 schließlich enthält ein kurzes Fazit.

2 User-Generated Content

Die Wikipedia (Stand: 03.09.2007) definiert den Begriff „User-Generated Content“ unter [1] wie folgt: „Inhalte (Content) in Medien wie Bilder und Texte, auch Audiodaten oder Filme, die ein oder mehrere Benutzer (User) selbst erzeugen (generieren). Diese Inhalte stehen anderen Interessierten – in der Regel kostenfrei – zur Verfügung.“

Erscheint diese Definition auf den ersten Blick noch relativ schlüssig, wird bei genauerer Untersuchung schnell klar, dass sie zahlreiche Unschärfen in Bezug auf mehrere Aspekte aufweist.

Zunächst wird der Ausdruck „User-Generated Content“ in der Regel lediglich in Bezug auf das Medium Internet angewendet. So fallen beispielsweise Leserbriefe in Printmedien eher nicht in den Bereich des UGC, obwohl auch sie regelmäßig die oben genannten Kriterien erfüllen.

Andererseits ist UGC – wie in der obigen Definition bereits angedeutet – nicht auf eine bestimmte „Inhaltsart“ festgelegt; sowohl Texte als auch Bilder, Grafiken, Animationen, Videos oder Musik können als UGC gelten. Streng genommen gibt es auch keinen Grund, Software von der Definition des UGC auszunehmen, da auch z. B. von Internetusern geschriebene Computerprogramme, die im Internet heruntergeladen werden können, alle oben beschriebenen Kriterien erfüllen, und zwar unabhängig davon, ob sie im Quellcode (Open Source-Software)

oder nur als vorkompiliertes Programm vorliegen. Gleiches gilt auch für „normale“ private Internetseiten.

Letztlich könnte folglich jede Form elektronischer Daten, die von Internetusern selbst erstellt wurde und anderen über das Internet zum Abruf zur Verfügung gestellt wird, als UGC angesehen werden.

Eine weitere Unschärfe des Begriffes „User-Generated Content“ bildet das Wort „User“. Im Zusammenhang mit UGC können oft drei Rollen unterschieden werden:

- Plattformbetreiber: Stellen die technische Infrastruktur (Server, Netzwerk, Datenbanken, Plattformsoftware etc.) bereit und betreiben diese;
- Verfasser: Stellen Inhalte zur Verfügung, müssen nicht immer identisch mit dem Urheber nach Urheberrechtsgesetz sein;
- Konsumenten: Rufen Inhalte ab.

Bei vielen typischen Projekten und Angeboten im Bereich des UGC übernehmen Unternehmen oder andere Organisationen die Rolle des Plattformbetreibers. Der Begriff „User“ dient daher häufig zur Abgrenzung einzelner privater Internetnutzer (meist Verfasser und Konsumenten) von diesen Organisationen.

Gerade durch die neuen technischen Möglichkeiten des Internets und des Web 2.0 ist jedoch nicht nur die früher strengere Trennung zwischen Inhalteanbieter und -nutzer, sondern auch die zwischen Plattformbetreiber eines Mediums (z. B. Buchverlag) und den Nutzern (Autor und Leser) stark aufgeweicht worden. Als Beispiel mag hier ein privater Forenbetreiber dienen, der auf einem angemieteten Server eine Forensoftware administriert und auch selbst in dem Forum aktiv ist. Dieser „User“ übernimmt gleichzeitig die Rollen des Plattformbetreibers, des Verfassers und des Konsumenten, z. B. beim Lesen fremder Forenbeiträge. Außerdem nutzen auch immer mehr Firmen typische Medien des UGC zu PR-Zwecken (z. B. Blogs von Vorstandsmitgliedern).

Gelegentlich wird der Begriff UGC sogar als beleidigend angesehen, weil er einen grundsätzlichen qualitativen Unterschied zwischen unentgeltlich zur Verfügung gestellten und kommerziellen Inhalten suggeriere, der aber nicht automatisch gegeben sei (vgl. [Him07]).

Kritisch angemerkt werden muss in jedem Fall, dass die angebotenen Inhalte auf vielen Plattformen, die sich selbst als UGC-Anbieter verstehen (z. B. YouTube, Clipfish.de), nicht wirklich UGC im eigentlichen Sinne darstellen. Dies gilt insbesondere für Ausschnitte aus Fernsehsendungen, Filmen sowie anderem urheberrechtlich geschützten Material, welche zwar von Internetnutzern auf die Plattform hochgeladen, jedoch nicht originär von diesen erstellt wurden.

Doch auch hier ist eine Abgrenzung nicht immer leicht möglich. Während bloße Ausschnitte aus Fernsehsendungen kaum als UGC angesehen werden können, stellen z. B. von Fans neu vertonte Szenen bereits einen Grenzfall dar. Ein weiteres Beispiel sind bewegte Foto-Collagen, bei denen im Internet verfügbare Fotos (z. B. von bekannten Personen) zu einem Video zusammengestellt und ggf. mit Hintergrundmusik hinterlegt werden. Auch wenn diese Beispiele aus urheberrechtlicher Sicht als bedenklich angesehen werden müssen (vgl. Abschnitt 4.3), kann

doch nicht negiert werden, dass hier auf der Basis bestehender Inhalte durch Komposition ein neues „Werk" entsteht, das durchaus als „User-generiert" angesehen werden kann.

Insgesamt lässt sich also festhalten, dass der Begriff „User-Generated Content" – wie letztlich auch der Begriff „Web 2.0" – nicht fest und trennscharf definiert ist. Eine wirklich treffende Präzisierung dürfte aufgrund der o. g. Schwierigkeiten auch kaum möglich sein. Insofern bleibt an dieser Stelle nur, auf den allgemeinen Sprachgebrauch zu verweisen, der durch die Wikipedia-Definition recht gut wiedergegeben wird.

3 Qualität

Zum Begriff „Qualität" gibt es eine Vielzahl von Definitionen. Einige Beispiele sollen hier genannt werden:

- „Die Gesamtheit von Merkmalen einer Einheit bezüglich ihrer Eignung, festgelegte und vorausgesetzte Erfordernisse zu erfüllen" (DIN EN ISO 8402:1995-08, mittlerweile zurückgezogen).
- „Grad, in dem ein Satz inhärenter Merkmale Anforderungen erfüllt" (DIN EN ISO 9000: 2005).
- „Qualität soll darauf abzielen, die gegenwärtigen und zukünftigen Wünsche des Kunden zu erfüllen" (W. Edwards Deming).
- „Qualität ist die Einhaltung der Anforderungen" (Philip B. Crosby).
- „Übereinstimmung zwischen den festgestellten Eigenschaften und den vorher festgelegten Forderungen einer Betrachtungseinheit" (IEC 2371).

Gemeinsam ist allen diesen Definitionen, dass die Qualität mit der Erfüllung von Anforderungen zusammenhängt. Ein Objekt, das die an es gestellten Anforderungen in hohem Maße erfüllt, weist demzufolge eine hohe Qualität auf. Da jedoch sowohl Anforderungen als auch deren Erfüllung durch ein gegebenes Objekt in vielen Fällen nur subjektiv vorgegeben bzw. beurteilt werden können, ist eine objektive Qualitätsmessung oft unmöglich.

Dies gilt auch und ganz besonders für immaterielle Güter wie Medieninhalte und damit auch für UGC. Schon allein anhand der vielen verschiedenen Arten von UGC, die in Kapitel 2 genannt wurden, lässt sich ersehen, dass die Definition eines allgemein gültigen Anforderungskatalogs kaum sinnvoll ist. Hinzu kommt noch, dass verschiedene oder sogar derselbe Nutzer in unterschiedlichen Situationen und Nutzungskontexten vollkommen unterschiedliche Anforderungen an ein und denselben Inhalt haben können.

Einige typische Anforderungen an UGC wären etwa Korrektheit, Objektivität, Aktualität, Authentizität, Vollständigkeit, Verständlichkeit, Relevanz, Ästhetik und Unterhaltungswert. Wie bereits erwähnt, sind jedoch nicht alle diese Kriterien bei jeder Art von Content und in jeder Situation gleich wichtig. Während bei einem Enzyklopädie-Artikel oft großer Wert auf Korrektheit und Objektivität gelegt wird, steht bei einem Videoclip meist eher der Unterhaltungswert im Vordergrund.

Qualitätsmerkmale, die sich eher auf die Plattform als auf den UGC selbst beziehen, sind z. B. Ergonomie und Barrierefreiheit.

Eine weitere, extern vorgegebene Anforderung an UGC ist dessen Legalität, also die Übereinstimmung mit geltendem Recht. Von Bedeutung sind hier z. B. gesetzliche Regelungen des Urheberrechts und des Jugendschutzes, aber auch andere strafrechtliche Normen von Beleidigungen bis hin zur Volksverhetzung.

Abschließend bleibt festzuhalten, dass sich die Qualität von UGC in vielen Fällen nicht messen oder anderweitig objektiv und allgemein gültig beurteilen lässt. Abgesehen davon, dass verschiedene User völlig unterschiedliche Anforderungen an Inhalte haben können, ist auch der Erfüllungsgrad bezüglich einer Anforderung oft subjektiv.

Nichtsdestotrotz hat die Qualität von UGC Auswirkungen, z. B. auf den Erfolg der Plattform, über die er abrufbar ist. Einige dieser Konsequenzen werden in Kapitel 4 behandelt.

4 Konsequenzen der Qualität von UGC

4.1 Allgemeine Konsequenzen

Die Qualität von UGC auf einer Plattform hat naturgemäß in erster Linie Konsequenzen für die direkt Beteiligten, nämlich Plattformbetreiber, Verfasser und Konsumenten.

Die Folgen von „schlechtem" UGC für Konsumenten sind in der Regel noch überschaubar. Ein Internetnutzer, der sich ein langweiliges Video ansieht oder einen unverständlichen Blog-Eintrag liest, wird sich evtl. über die verlorene Zeit ärgern und die betreffende Plattform nicht wieder besuchen; einen echten Schaden wird er im Regelfall jedoch kaum erleiden. Einschränkend kann hier lediglich angemerkt werden, dass einem User mit geringer Medienkompetenz, der sich blindlings auf die Korrektheit von Informationen im UGC verlässt, durchaus Nachteile entstehen können (z. B. bei Aktienkaufempfehlungen in einem Blog). Dies gilt jedoch in gleichem Maße auch für andere Medien und Inhalte und stellt keine Besonderheit von UGC dar.

Verfasser, die qualitativ minderwertigen Content einstellen, haben im Regelfall ebenfalls keine gravierenden Konsequenzen zu befürchten. Da sie normalerweise unentgeltlich tätig sind, entstehen ihnen bei „Schlechtleistung" keine finanziellen Nachteile, lediglich ein gewisser Vertrauens- und Ansehensverlust unter den Plattformnutzern oder im Extremfall die Sperrung des Benutzerkontos steht zu erwarten.

Von entscheidender Bedeutung ist die Content-Qualität hingegen für den Plattformbetreiber, da minderwertiger Content im Normalfall zumindest mittelfristig zu sinkenden Nutzerzahlen führt. Handelt es sich bei der Plattform nicht um ein rein privates Angebot, aus dem der Betreiber keinerlei materiellen Nutzen zieht, mag dies noch zu verschmerzen sein. Für kommerzielle Anbieter aber bedeuten fallende Nutzerzahlen meist auch sinkende Einnahmen, z. B. aus Werbung auf der Plattform. Es ist daher meist ein zentrales Anliegen des Betreibers, die Qualität des Contents auf seiner Plattform möglichst hoch zu halten, und sei es nur aus schierem Idealismus. Hierzu kann er sich verschiedener Qualitätssicherungs- und Anreizsysteme bedienen, von denen einige in Kapitel 5 exemplarisch beschrieben werden.

4.2 Gesellschaftliche Konsequenzen

In den vergangenen Jahren hat UGC immer weiter an Bedeutung gewonnen. Viele Internetuser nutzen heute z. B. Onlinenachschlagewerke, lassen sich durch Videos auf UGC-Plattformen unterhalten und beziehen Informationen aus Blogs, die früher – wenn überhaupt – nur in Zeitungen und Zeitschriften verfügbar waren. Hieraus resultiert, dass die Qualität von UGC zunehmend auch gesamtgesellschaftliche Konsequenzen haben kann.

Als Beispiel sei hier ein Video genannt, das im März 2007 angeblich von einem griechischen Jugendlichen auf einer UGC-Plattform veröffentlicht wurde und in welchem Kemal Atatürk, Begründer der modernen Türkei, als homosexuell bezeichnet wird. Die Veröffentlichung des Videos löste in der Türkei einen Sturm der Entrüstung aus; angeblich türkische User reagierten mit wütenden Kommentaren und Videos, in denen im Gegenzug Griechenland herabgewürdigt werden sollte. Aufgrund des Videos wurde sogar die Türk Telekom per einstweiligen Verfügung eines Istanbuler Gerichtes gezwungen, den Zugriff auf die betreffende Videoplattform zu sperren. Dieser Vorfall zeigt, dass UGC mittlerweile schnell eine große Zahl von Konsumenten erreichen und daher eine große gesellschaftliche Bedeutung erlangen kann.

Psychologische Studien (vgl. [LSO+03]) belegen darüber hinaus, dass Menschen bevorzugt diejenigen Informationen im Gedächtnis bleiben, die ihnen zuerst präsentiert werden und von deren Richtigkeit sie überzeugt sind. Dies gilt sogar dann, wenn ihnen im Nachhinein andere Informationen gegeben werden, aus denen hervorgeht, dass die ursprünglichen falsch waren.

Vor diesem Hintergrund zeigt sich klar die große Verantwortung der Verfasser und Plattformbetreiber für die Qualität des veröffentlichten UGC, da eine schlechte Qualität nicht nur negative Konsequenzen für sie selbst, sondern auch für eine breitere Öffentlichkeit haben kann. Zwar gilt dies grundsätzlich gleichermaßen auch für sonstige Inhalte in beliebigen Medien. Es kann jedoch bezweifelt werden, ob z. B. Jugendliche, die im Internet in ihrer Freizeit Videos zu einem Thema veröffentlichen, sich dieser Verantwortung ebenso bewusst sind und die Folgen ihrer Handlungen ebenso gut abschätzen können wie etwa ein ausgebildeter Journalist, der für eine Tageszeitung einen Artikel zum selben Thema verfasst.

Somit könnten also allenfalls die Plattformbetreiber bei fragwürdigen Inhalten korrigierend eingreifen, wodurch sie jedoch in das Spannungsfeld zwischen Meinungsfreiheit und verschiedenen Kulturen sowie Rechtssystemen geraten.

4.3 Juristische Konsequenzen

Nicht oder nur sehr eingeschränkt gelten die in Abschnitt 4.1 getroffenen Aussagen im Bezug auf die Anforderung der Legalität von UGC; dieses Kriterium nimmt daher eine Sonderstellung ein.

Für User ist es noch relativ einfach, sich gegen straf- oder zivilrechtliche Verfolgung (Unterlassungs- oder Schadensersatzansprüche) wegen entsprechender Inhalte zu schützen, indem sie darauf verzichten, solche zu veröffentlichen bzw. zu konsumieren. Ein Restrisiko resultiert jedoch aus unterschiedlichen Rechtsauffassungen über die Rechtswidrigkeit der Inhalte. Dies kann sowohl Privatpersonen oder -organisationen betreffen, die (zumindest angeblich) ihre

Rechte verletzt sehen, als auch staatliche Strafverfolger, welche die Inhalte für rechtswidrig halten (Beispiel: Strafverfahren gegen Holger Voss, kurze Zusammenfassung in [2]). Fast vollständigen Schutz bietet Usern daher nur die Nutzung von technischen Anonymisierungsdiensten, wodurch ein Rückschluss auf eine konkrete Person unmöglich wird.

Ungleich schwieriger ist die Situation für Plattformbetreiber. In der Regel haben sie nur sehr eingeschränkte Einflussmöglichkeiten auf die Inhalte. Eine manuelle Vorabprüfung aller veröffentlichten Inhalte ist meist schon allein aufgrund der Menge des Materials unmöglich; automatische Filter (z. B. zur Sperrung urheberrechtlich geschützten Materials) werden zwar immer wieder gefordert und angekündigt, sind jedoch technisch bisher nicht umsetzbar.

Trotzdem haften Plattformbetreiber nach deutschem Recht ggf. als so genannte Mitstörer, wenn auf den von ihnen betriebenen Plattformen rechtswidrige Inhalte veröffentlicht werden. In welchen Fällen eine solche Mitstörerhaftung möglich ist, ist in Deutschland noch nicht höchstrichterlich entschieden. Es gibt zwar schon mehrere Urteile zu diesem Thema, die sich jedoch teilweise widersprechen. Unbestritten ist allgemein, dass Plattformbetreiber haften, wenn sie rechtswidrige Inhalte nicht unverzüglich löschen, nachdem sie ihnen zur Kenntnis gebracht wurden. Dies versetzt die Betreiber jedoch in die unangenehme Lage, selbst bewerten zu müssen, ob ein Inhalt tatsächlich rechtswidrig ist. Sperren sie den Zugriff, wird ihnen vonseiten der Verfasser und Konsumenten oft Zensur vorgeworfen, tun sie dies nicht, riskieren sie straf- und zivilrechtliche Konsequenzen.

Schließlich wird diese Situation noch weiter dadurch verkompliziert, dass bestimmte Inhalte in verschiedenen Rechtssystemen unterschiedlich bewertet werden. Während etwa rechtsradikale Propagandavideos in Deutschland rechtswidrig sind, werden sie in den USA vom Recht auf freie Meinungsäußerung gedeckt. Gerade für international tätige Plattformbetreiber erschwert dies die Entscheidung über eine Sperrung oder Löschung von Inhalten zusätzlich.

5 Beispiele

5.1 Wikipedia

Das sehr bekannte Wikipedia-Projekt hat sich zum Ziel gesetzt, mithilfe Freiwilliger eine für den Nutzer kostenlose, möglichst umfassende Enzyklopädie aufzubauen. Naturgemäß werden an ein solches Projekt hohe Qualitätsanforderungen gestellt, insbesondere hinsichtlich Korrektheit, Objektivität, Vollständigkeit und Verständlichkeit.

Da grundsätzlich jeder Internetnutzer – unabhängig von seiner Qualifikation in einem bestimmten Wissensgebiet – neue Wikipedia-Artikel erstellen und bestehende ändern kann, stellte sich bereits früh die Frage, ob und wie diese Anforderungen erfüllt werden können. Schließlich ist nicht sichergestellt, dass die Änderung des Artikels zu einer Qualitätssteigerung führt, ebenso ist auch eine Verschlechterung des Artikels möglich. Kritiker, wie z. B. Robert Dale McHenry (ehemaliger Chefredakteur und Vizepräsident der Encyclopaedia Britannica), äußerten die Ansicht, dass gerade diese Offenheit der Wikipedia gegenüber Änderungen (unabhängig davon, ob sie einen Artikel verbessern oder verschlechtern) nach den Prinzipien der Wahrscheinlich-

keitsrechnung langfristig zu einem Einpendeln der Artikelqualität auf ein Mittelmaß führen werde [3].

Außerdem war und ist die Wikipedia auch immer wieder gezielten Angriffen ausgesetzt. Neben schierem Vandalismus sind auch Fälle publik geworden, in denen gezielt Wikipedia-Artikel verändert wurden, um Lügen über politische Gegner zu verbreiten oder Firmen oder Privatpersonen anderweitig in Misskredit zu bringen (vgl. z. B. [4]). Gerade bei Artikeln zu kontroversen Themen kommt es gelegentlich zu sog. Editwars, bei denen Vertreter verschiedener Meinungen immer abwechselnd versuchen, die jeweils eigene Sichtweise in den Artikel einzuarbeiten und die konträre zu löschen.

Aus diesen Gründen wurde in den vergangenen Jahren eine ganze Reihe von Mechanismen und Hilfsmitteln erarbeitet, um hochqualitative Wikipedia-Artikel zu kennzeichnen und zu erhalten, Nutzer zur Überarbeitung verbesserungswürdiger Artikel aufzurufen sowie böswillige Angriffe zu erkennen und schnell rückgängig machen zu können. Einige Beispiele sollen hier stichpunktartig genannt werden:

- Versionshistorie: Über ein Versionierungssystem sind alte Versionen eines jeden Artikels abrufbar und können – z. B. bei Fällen von Vandalismus – schnell wieder hergestellt werden (Beispiel unter [5]).
- Lesenswerte und exzellente Artikel: Auszeichnung, die für besonders hochqualitative Artikel nach einem Peer Review vergeben wird, um Nutzer auf diese Inhalte aufmerksam zu machen und einen Anreiz zur weiteren Verbesserung bestehender Artikel zu schaffen (Liste „exzellenter Artikel“ unter [6]).
- Bewertungsbausteine: Können von jedem Nutzer in Artikel eingefügt werden, um auf einen konkreten Überarbeitungsbedarf des Artikels (z. B. „Belege fehlen“) hinzuweisen, Benutzerdokumentation ist unter [7] abrufbar.
- Geschützte Seiten: Artikel können von Administratoren zeitweilig gegen Bearbeitung geschützt werden, um Vandalismus und Editwars zu begegnen; weitere Erläuterungen finden sich unter [8].
- CryptoDerk's Vandal Fighter (CDVF): Open Source-Tool zur Vandalismusbekämpfung, welches Änderungen an Artikeln überwacht und dem Benutzer in Echtzeit meldet. Detailliertere Informationen sind unter [9] abrufbar.
- Wikipedia-Scanner: Auf [10] zur Verfügung gestellte Anwendung zur leichteren Erkennung gezielter Änderungen durch Interessentengruppen. Dazu werden die in der Wikipedia erfassten IP-Adressen, von denen aus anonyme Änderungen vorgenommen wurden, mit den fest vergebenen IP-Bereichen größerer Organisationen (z. B. Ministerien, Unternehmen, Sekten) verglichen, sodass erkennbar wird, aus wessen Netz heraus die Änderung vorgenommen wurde.

Obwohl immer wieder Fehler in der Wikipedia entdeckt und veröffentlicht werden, scheinen die genannten Maßnahmen trotzdem insgesamt wirksam zu sein. In einer Studie verglich das renommierte Wissenschaftsmagazin *Nature* die Qualität von Wikipedia und Encyclopaedia Britannica anhand 42 ausgewählter Artikel [Gil05]. Dabei wurden in beiden Enzyklopädien gleich

viele schwerwiegende Fehler gefunden; lediglich bezüglich kleinerer Fehler schnitt die Encyclopaedia Britannica besser ab:

	Wikipedia	Encyclopaedia Britannica
Schwerwiegende Fehler	4	4
Sachliche Fehler, Auslassungen, missverständliche Formulierungen	162	123

Abbildung 1: Fehleranzahl in Wikipedia und Encyclopaedia Britannica

Die Studie kommt folglich zu dem Ergebnis, dass die Wikipedia zwar wie erwartet Fehler aufweist, die Content-Qualität jedoch insgesamt mit der der kommerziellen Encyclopaedia Britannica durchaus vergleichbar ist.

5.2 YouTube

YouTube ist eine Videoplattform im Besitz der Google Inc., auf der kostenlos von anderen Nutzern eingestellte Videodateien abgerufen werden können.

Naturgemäß werden an Videos auf YouTube vollkommen andere Anforderungen gestellt als an Wikipedia-Artikel. In den meisten Fällen steht hier der Unterhaltungswert im Vordergrund, bei einigen Videos evtl. auch der Informationsgehalt.

Obwohl der Unterhaltungswert kaum objektivierbar ist, bietet YouTube den Konsumenten die Möglichkeit, Inhalte auf einer Skala mit 1 („Poor“) bis 5 („Awesome!“) Sternen zu bewerten. Ein Durchschnittswert der abgegebenen Bewertungen wird zu jedem Video angezeigt. Außerdem können Konsumenten Textkommentare zu den Inhalten abgeben, die unterhalb des Videos angezeigt werden. Um dem Nutzer schnellen Zugriff auf (zumindest potenziell) interessante Inhalte zu gewähren, kann dieser sich Listen z. B. der meistabgerufenen, am besten bewerteten oder meistkommentierten Videos anzeigen lassen. Ebenso gibt es eine Liste der Videos, die andere User am häufigsten zu ihren persönlichen Favoriten hinzugefügt haben. Ziel ist es hier also nicht erster Linie, Content mit schlechter Qualität zu löschen, sondern die Konsumenten auf hochqualitative Inhalte hinzuweisen und diese leicht zugänglich zu machen.

Diese Maßnahmen scheinen Erfolg zu haben, da YouTube laut [11] mit deutlichem Vorsprung Marktführer auf dem US-Onlinevideomarkt ist:

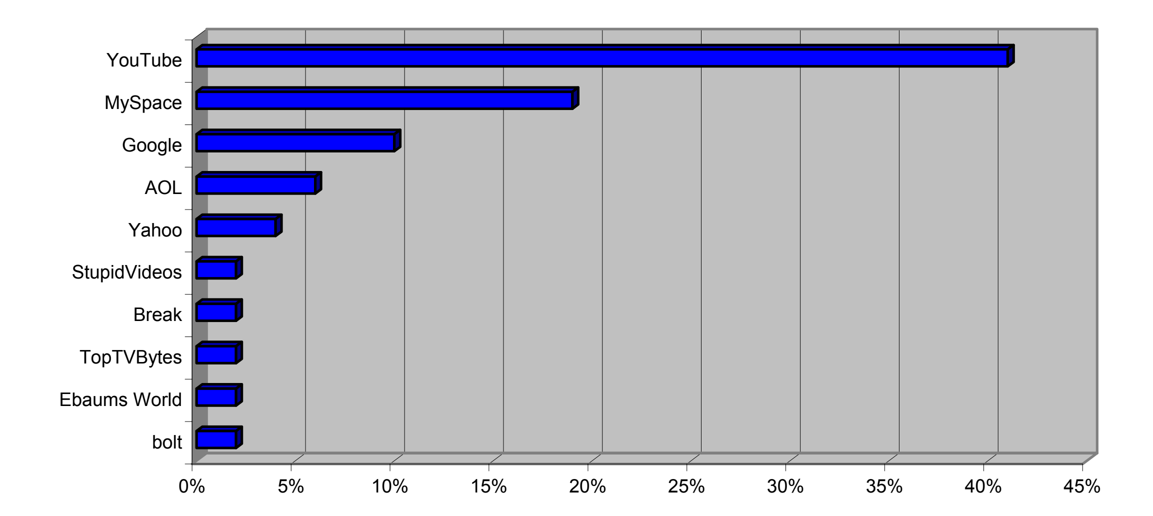


Abbildung 2: Marktanteile unter den Top-20-Onlinevideoanbietern (Dez. 06)

Der Anteil der vollkommen originär vom Verfasser erstellten Inhalte ist auf YouTube im Vergleich zu anderen UGC-Plattformen relativ gering. Dies bringt – wie in Abschnitt 4.3 bereits beschrieben – juristische Probleme, wie etwa Unterlassungsansprüche seitens der Rechteinhaber, mit sich. Außerdem muss sich YouTube/Google zurzeit einer Schadensersatzklage des amerikanischen Medienunternehmens Viacom erwehren (vgl. z. B. [12]). Auch Gerichtsurteile, die den Zugriff auf die Plattform wegen rechtswidriger Inhalte unterbinden lassen (vgl. Abschnitt 4.2), mindern zumindest potenziell die Besucherzahlen und damit eventuelle Werbeeinnahmen.

Die Sperrung bzw. Löschung rechtswidriger Inhalte kann bisher nur durch manuelle Eingriffe geschehen. Daher wäre ein „Copyright-Filter", der urheberrechtlich geschütztes Material automatisch ausfiltert, aus Sicht der Plattformbetreiber wünschenswert, seine technische Umsetzbarkeit erscheint momentan jedoch unwahrscheinlich. Nach eigenen Aussagen arbeitet YouTube zurzeit an einem System, das Inhalte über Hashwerte identifizieren und so zumindest das wiederholte Einstellen einmal gelöschter Inhalte unterbinden soll (vgl. z. B. [13]). Ob diese Bemühungen letztlich von Erfolg gekrönt sein werden, scheint ebenfalls fraglich. So müsste durch ein solches System z. B. verhindert werden, dass schon kleinste Änderungen am Videomaterial zu einem anderen Hashwert führen, wodurch das Video nicht wiedererkannt würde. Das Problem rechtswidriger Inhalte auf YouTube muss daher als weitestgehend ungelöst angesehen werden.

5.3 Gloob.tv

Gloob.tv stellt einen werbefinanzierten Mehrwertdienst dar, der sich gerade die gemischte Content-Qualität auf anderen Plattformen zunutze macht. Dazu wird hier kaum eigener Content angeboten, stattdessen sichten die Mitarbeiter von Gloob.tv die Inhalte auf anderen Onlinevideoplattformen und stellen die ihrer Meinung nach besten Inhalte auf der eigenen Webpräsens nach Kategorien sortiert zusammen. Der Konsument erhält dadurch eine vergleichsweise kleine, aber hochqualitative Menge an „hand-picked" Content, ohne sich selbst auf die Suche nach entsprechenden Inhalten machen zu müssen.

Das Geschäftsmodell von Gloob.tv stellt ein interessantes Beispiel für Dienste dar, die zwar auf bestehendem UGC aufsetzen, ohne jedoch selbst als eigentliche UGC-Plattform aufzutreten. Gerade die gemischte Content-Qualität auf den bestehenden UGC-Plattformen ist dabei eine entscheidende Grundlage für den Erfolg des Geschäftsmodells. Wäre die Content-Qualität auch dort gleichbleibend hoch, gäbe es für die Konsumenten keinen Anreiz, den Mehrwertdienst zu nutzen.

6 Fazit

Wie auch „Web 2.0" ist der Begriff des „User-Generated Content" letztlich nicht trennscharf definierbar, wird jedoch im allgemeinen Sprachgebrauch für Inhalte im Internet benutzt, die von Usern zumindest teilweise selbst erstellt und anderen meist unentgeltlich zur Verfügung gestellt werden. Auch die Beurteilung der Qualität der bereitgestellten Inhalte ist oft nicht objektiv möglich, sondern lediglich subjektiv, abhängig von den Anforderungen und Vorlieben des konkreten Konsumenten.

Nichtsdestotrotz hat die Content-Qualität große Bedeutung, insbesondere für die Plattformbetreiber. Während Verfasser und Konsumenten nur im Einzelfall (z. B. bei Nichterfüllung der Anforderung „Legalität") ernsthafte Konsequenzen von niedriger Content-Qualität zu befürchten haben, hängt für kommerzielle Plattformbetreiber oft der Erfolg ihres Geschäftsmodells von der Content-Qualität ab. Auch im nicht-kommerziellen Bereich wird eine hohe Content-Qualität vom Plattformbetreiber meist allein schon aus Idealismus angestrebt.

Wie in Kapitel 5 dargestellt, haben verschiedene Plattformbetreiber daher unterschiedliche technische und organisatorische Maßnahmen ergriffen, um die Qualität des UGC auf ihrer Plattform sicherzustellen bzw. zu erhöhen. Zum Teil waren diese Maßnahmen bereits von Erfolg gekrönt, einige Qualitätsprobleme – insbesondere bezüglich Legalität und Authentizität – müssen jedoch noch als weitgehend ungelöst angesehen werden.

Schließlich konnte anhand des Beispiels Gloob.tv aufgezeigt werden, dass gerade die gemischte Content-Qualität auf vielen UGC-Plattformen auch Möglichkeiten zur Etablierung neuer Geschäftsmodelle bieten kann.

Literaturverzeichnis

[Gil05] J. Giles: „Internet encyclopaedias go head to head", Nature 438, pp. 900-901, 2005.

[Him07] G. Himmelein: Hin, her und wieder zurück – Erster weltweiter Copyright-Gipfel der Verwertungsgesellschaften, c't, magazin für computertechnik 13, pp. 34-35, 2007.

[LSO+03] S. Lewandowsky, W. Stritzke, K. Oberauer, M. Morales: Memory for Fact, Fiction, and Misinformation, Psychological Science 16 (3), pp 190-195, 2003.

Verzeichnis von Webadressen

[1] Wikipedia-Artikel zu „User Generated Content“, http://de.wikipedia.org/wiki/User_Generated_Content. Zuletzt abgerufen 03.09.2007.

[2] Telepolis-Artikel „Eine deutsche Justizposse“, http://www.heise.de/tp/r4/artikel/13/13919/1.html. Zuletzt abgerufen 04.09.2007.

[3] Robert McHenry: The Faith-Based Encyclopedia, http://www.techcentralstation.com/111504A.html. Zuletzt abgerufen 04.09.2007.

[4] Artikel des „New Yorker“ zur böswilligen Veränderung von Wikipedia-Artikeln im US-Senatswahlkampf, http://www.newyorker.com/archive/2006/11/06/061106ta_talk_paumgarten. Zuletzt abgerufen 04.09.2007.

[5] Versionshistorie zum Wikipedia-Artikel zu „Wikipedia“, http://de.wikipedia.org/w/index.php?title=Wikipedia&action=history. Zuletzt abgerufen 04.09.2007.

[6] Liste „exzellenter“ Wikipedia-Artikel, http://de.wikipedia.org/wiki/Liste_exzellenter_Artikel. Zuletzt abgerufen 04.09.2007.

[7] Dokumentation zu Bewertungsbausteinen der Wikipedia, http://de.wikipedia.org/wiki/Wikipedia:Bewertungsbausteine. Zuletzt abgerufen 04.09.2007.

[8] Erläuterung zur Wikipedia-Funktion der geschützten Seiten, http://de.wikipedia.org/wiki/Wikipedia:Gesch %C3 %BCtzte_Seiten. Zuletzt abgerufen 04.09.2007.

[9] Erläuterungen zu CryptoDerk's Vandal Fighter im User-Profil von „CryptoDerk“, http://en.wikipedia.org/wiki/User:CryptoDerk/CDVF. Zuletzt abgerufen 04.09.2007.

[10] Wikipedia-Scanner, http://wikiscanner.virgil.gr. Zuletzt abgerufen 04.09.2007.

[11] Compete.com zum Marktanteil von YouTube, http://blog.compete.com/2007/02/02/google-video-youtube-top-video-sites/. Zuletzt abgerufen 04.09.2007.

[12] Bericht zum Rechtsstreit zwischen Viacom und YouTube/Google auf Groklaw.net, http://www.groklaw.net/article.php?story=20070510132915305&query=Google+Viacom. Zuletzt abgerufen 04.09.2007.

[13] Eintrag zu YouTubes geplantem Hash-Filter aus dem GoogleWatchBlog, http://www.googlewatchblog.de/2007/07/31/youtube-filter-startet-im-september/. Zuletzt abgerufen 04.09.2007.

Data Mining im User-Generated Content

Nico Albrecht

1 Einleitung

Internetnutzer stellen im Zuge der Web 2.0-Bewegung zunehmend eigene Webinhalte zur Verfügung. Diese Daten eignen sich in vielen Fällen zur automatisierten Auswertung, im Zuge derer wertvolle Informationen gewonnen werden können. Dieser Beitrag widmet sich der Beantwortung der Fragen, inwiefern nutzergenerierte Inhalte wirtschaftlich sinnvoll analysiert werden können und welche zusätzlichen Möglichkeiten sich Nutzern bieten, die von ihnen oder von anderen bereitgestellten Inhalte zu verwerten. Dazu werden nach einer Definition der Begriffe Data Mining und User-Generated Content im zweiten Kapitel Fallbeispiele vorgestellt, in denen bereits Techniken des Data Minings in nutzergenerierten Inhalten angewendet werden. Im Anschluss daran werden in Kapitel 4 Einsatzpotenziale identifiziert, wie durch Data Mining-Verfahren aus User-Generated Content Informationen generiert werden können. Dies geschieht zum einen aus der Sicht von Unternehmen und zum anderen aus Nutzersicht. Die Arbeit schließt mit einer Zusammenfassung und gibt einen Ausblick auf künftige Entwicklungen, die eine breitere Diffusion der behandelten Fragestellung ermöglichen.

2 Begriffliche Grundlagen

2.1 Data Mining

Im Zuge einer ständigen Zunahme von relativ leicht zugänglichen Daten durch unternehmensinterne (z. B. Enterprise-Resource-Planning- oder Supply-Chain-Management-Anwendungen) und unternehmensexterne (z. B. das Internet) Systeme wächst gleichzeitig der Bedarf an Methoden, diese Datenmenge zu strukturieren und Informationen hieraus zu generieren. Insbesondere die Digitalisierung von Daten und die Homogenisierung von Systemen vereinfachen die Datenanalyse und steigern somit den Bedarf, potenziell in einem Datenbestand verborgene Informationen zu extrahieren. Während es bei einer relativ geringen Menge an Daten noch einfach erscheint, Regelmäßigkeiten zu erkennen, stellen insbesondere größere Datenbestände den Anwender vor das Problem, automatisierte Verfahren einzusetzen, um in akzeptabler Zeit Muster erkennen zu können und somit Informationen bzw. Entscheidungsunterstützungen zu generieren.

Das Forschungsfeld der statistischen Datenanalyse bietet bereits eine Vielzahl von Werkzeugen zur Analyse von strukturiert vorliegenden Daten. Der Fokus hierbei hat sich in den letzten Jahren von einer automatisierten Überprüfung zu einer automatisierten Generierung von Hypo-

thesen verschoben.[1] Die Interdisziplinarität der sich hieraus ergebenen Forschungsrichtung des Data Minings zeigt sich in der Vielzahl der Anwendungsgebiete.[2]

Eine ausführliche Diskussion der Begriffe Data Mining und Knowledge discovery in databases liefert BENSBERG, dessen Definition von Data Mining als „integrierte[n] Prozess […], der durch Anwendung von Methoden auf einen Datenbestand Muster entdeckt"[3] dieser Arbeit zugrunde gelegt wird.[4] Angelehnt an FAYYAD, PIATETSKY-SHAPIRO und SMYTH wird im Folgenden ein kurzer Überblick über die wichtigsten Verfahren des Data Minings gegeben.[5] Die Autoren schlagen eine Systematisierung nach Aufgaben des Data Minings vor:

- Klassifikation (classification),
- Regression (regression),
- Segmentierung (clustering) und
- Abhängigkeitsanalyse (dependency modeling).

Zu Methoden der *Klassifikation* sind die Diskriminanzanalyse, die K-Nearest-Neighbour-Methode, die Entscheidungsbauminduktion, die Regelinduktion und Künstliche Neuronale Netze (KNN) zu zählen. Diese Verfahren zielen darauf ab, anhand von Merkmalen bestimmter Objekte letztere in Klassen zu unterteilen.

Bei den *Regressionsverfahren* werden Beziehungen zwischen einer abhängigen Variable (Zielvariable) und einer oder mehreren unabhängigen Variablen gesucht. Falls Zusammenhänge mit einer ausreichenden statistischen Signifikanz bestehen, können Niveaus der Zielvariablen prognostiziert werden. Zu den Regressionsverfahren gehören neben der multivariaten linearen Regression auch Künstliche Neuronale Netze.

Gegenstand der *Segmentierung* ist die Identifikation von möglichst homogenen Segmenten aus einer ungeordneten Menge von Objekten. Im Unterschied zur Klassifikation werden die Klasseneinteilungen automatisiert ermittelt. Methoden der Segmentierung umfassen Kohonen-Netze (selbstorganisierende Karten) und Verfahren des maschinellen Lernens.

Verfahren der *Abhängigkeitsanalyse* suchen und analysieren Dependenzen zwischen Attributen bestimmter Objekte. Beispiele für Methoden der Abhängigkeitsanalyse sind neben Regressions- und Entscheidungsbaumverfahren die Assoziations- oder Warenkorbanalyse.

---

[1] Vgl. Fayyad, U. M. (1998), S. 40 f.

[2] Eine Übersicht über unterschiedliche Anwendungsbereiche findet sich bei Hagedorn, J., Bissantz, N., Mertens, P. (1997).

[3] Bensberg, F. (2001), S. 64.

[4] Diese Definition geht auf Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. (1996), S. 6 zurück, die Knowledge Discovery in Databases (KDD) als „the non trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" definieren. Die Begriffe Data Mining und KDD werden häufig synonym verwendet.

[5] Vgl. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. (1996d), S. 85.

2.2 User-Generated Content

Das Nutzungsverhalten vieler Anwender im Internet weist in den letzten Jahren eine starke Tendenz zu aktiver Teilnahme anstatt ausschließlicher Rezeption der Inhalte auf. Begünstigende Faktoren hierfür sind unter anderem technische Veränderungen, wie eine zunehmende Verbreitung breitbandiger und kostengünstiger Internetverbindungen sowie die vereinfachte Usability von quelloffener Software. Auch der andauernde Erfolg von Open Source-Projekten, wie Wikipedia, Ubuntu oder das Project Gutenberg, die einen Teil des partizipativen Webs im Rahmen des Web 2.0-Gedankens[6] darstellen, tragen dazu bei, dass die Intensität der aktiven Teilnahme von nicht-professionellen Nutzern sowie deren Anzahl beständig zunimmt. So registriert beispielsweise der Webdienst Technorati eine stark steigende Anzahl von Blogs und damit verbunden eine Zunahme der Vernetzung durch Trackbacks und Kommentare.[7]

In der wissenschaftlichen Literatur findet sich zu dem Begriff User-Generated Content häufig der Verweis auf eine OECD-Studie zum partizipativen Web.[8] In dieser werden für User-Generated Content drei konstituierende Merkmale genannt: publication requirement, creative effort und creation outside of professional routines. Nutzergenerierte Inhalte müssen der OECD-Definition zufolge zunächst in nicht-flüchtiger Form im Internet publiziert werden, Chatnachrichten oder E-Mails zählen demnach nicht dazu. Dennoch fallen explizit auch Inhalte unter diese Definition, die auf nicht-öffentlich zugängigen Seiten gespeichert sind; hierzu zählen beispielsweise Einträge in geschlossenen Diskussionsforen oder nicht-öffentliche Profile aus Social-Networking-Sites. Zudem wird für User-Generated Content ein Mindestmaß an kreativem Schaffen gefordert, das über eine bloße Umordnung, Dateitransformation oder Sammlung von nicht-selbsterstellten Inhalten hinausgeht. Durch den Term „User" wird angedeutet, dass die Inhalte von nicht-professionellen Nutzern generiert werden. User-Generated Content kann demnach nur dort erschaffen werden, wo Personen nicht im Rahmen einer festen Organisationsstruktur handeln. Für diese Arbeit werden der OECD-Studie folgend unter dem Begriff User-Generated Content im Internet publizierte Inhalte verstanden, die außerhalb eines professionellen Umfeldes mit einem Mindestmaß an Kreativität erstellt wurden. Zur Veranschaulichung werden in Abbildung 1 Beispiele für User-Generated Content gegeben.

Kategorie	Beispiel	Beispielwebseiten
Text	Tags	Metadaten zu Fotos auf www.flickr.com
	Bewertungen	Produktbewertungen auf www.kelkoo.de
	Rezensionen	Buchrezensionen auf www.Amazon.de
	Blogeinträge	Artikel des Bloggers auf www.basicthinking.de/blog/
	Diskussionsforeneinträge	Einträge auf forum.spiegel.de/
	Gästebucheinträge	Einträge auf www.berlin.de/buergeraktiv/gaestebuch/
	Kommentare	Texte anderer Nutzer auf www.basicthinking.de/blog
	Wiki-Einträge	Artikel auf de.wikipedia.org
	Profile in Social-Networking-Sites	Nutzerprofile auf www.xing.com

---

6 Zu den Grundüberlegungen des Web 2.0 und des partizipativen Webs vgl. O'Reilly, T. (2005) sowie Beck, A. (2007).

7 Die Anzahl der Blogs wurde im April 2007 mit 70 Mio. angegeben, vgl. Sifry, D. (2007).

8 Vgl. OECD 2007. Hier wird von „User-created Content" gesprochen. In diesem Beitrag werden die Begriffe User-Generated Content, User-Created Content und nutzergenerierte Inhalte synonym verwendet.

Kategorie	Beispiel	Beispielwebseiten
Grafikdateien	Fotos selbsterstellte Bilder 3D-Objekte	Fotos auf www.flickr.com Bilder auf www.deviantart.com In SecondLife erstellte 3D-Objekte
Videos	Video-Podcasts Eigenständige Videoclips	Videoclips auf ifranz.org Videoclips auf www.youtube.com
Audiodateien	Audio-Podcasts Hörbücher Musik	Audioclips auf http://www.spreeblick.com/ category/radio-radio/ Hörbücher auf vorleser-amateure.de.vu Selbst produzierte Musik auf www.myspace.com
Sonstiges	Nicht-textliche Bewertungen Geokoordinaten Social Bookmarks	Bewertungen auf www.youtube.com Koordinaten auf loc.alize.us Bookmarksammlungen auf del.icio.us

Abbildung 1: Beispiele für User-Generated Content

Im folgenden Kapitel werden reale Anwendungsfälle beschrieben, die einen Überblick über die Einsatzpotenziale geben, die Data Mining in User-Generated Content bietet.

3 State of the Art des Einsatzes von Data Mining in User-Generated Content

3.1 Tag Clouds

Als Tag Clouds wird eine Möglichkeit zur Visualisierung der Relevanz bestimmter Begriffe innerhalb einer oder mehrerer Texte bezeichnet.[9] Hierbei wird zunächst eine Liste von Textlinks, auch Tags genannt (z. B. zu Suchergebnissen zu einem bestimmten Begriff), generiert und alphabetisch geordnet. Anschließend wird die Schriftgröße der Links proportional zur Bedeutung bzw. Häufigkeit der Begriffe gesetzt, ggf. werden bedeutende Tags auch farblich hervorgehoben. Dies ergibt eine einfache, intuitiv begreifbare Möglichkeit der Darstellung von Relevanzunterschieden. In Abbildung 2 ist eine typische Tag Cloud der Webseite von Spiegel Online dargestellt.



Abbildung 2: Tag Cloud der Seite www.spiegel.de[10]

---

9 Vgl. Rivadeneira et al. (2007) bzw. Szugat, M., Gewehr, J. E., Lochmann, C. (2006).

10 Quelle: www.spiegel.de (2007).

Vielfach erfolgt die Gewichtung ausschließlich aufgrund der Anzahl der Einträge zu einer bestimmten Kategorie, wobei der Name der Kategorie zugleich den Tag bezeichnet. Liegt eine solche Strukturierung nicht vor und sind beispielsweise die zugrunde liegenden Texte bzw. Suchergebnisse nicht ausreichend kategorisiert, so können Verfahren der Klassifikation bei einer (semi-)automatisierten Verschlagwortung helfen.

3.2 Amazon-Empfehlungen

Ein weiteres Praxisbeispiel des Einsatzes von Data Mining in User-Generated Content sind die automatisiert generierten Empfehlungen von amazon.com.[11] Hierbei wird auf die Warenkorbanalyse,[12] ein Verfahren der Abhängigkeitsanalyse, zurückgegriffen. Amazon trackt zu diesem Zweck die Klickpfade sowie die Artikel, zu denen Bewertungen verfasst wurden. Mithilfe eines automatisierten Verfahrens wird ermittelt, welche Artikel(kategorien) oder Bewertungen mit einem bestimmten statistischen Signifikanzniveau im Zusammenhang mit anderen Artikel(kategorien) bzw. Bewertungen angesehen, gekauft oder verfasst wurden. Als Ergebnis liefert das Verfahren Aussagen nach dem Schema „x % der Nutzer, die eine positive Bewertung des Artikels A verfasst haben, haben auch Artikel B positiv beurteilt“. Diese Aussagen können im Zusammenhang mit personalisierten Werbemaßnahmen verwendet werden, z. B. als Informationen auf einer Seite oder als E-Mail, in der auf für einen bestimmten Nutzer möglicherweise interessante Artikel hingewiesen wird. Als vorteilhaft aus Sicht von Amazon erweist sich hierbei, dass die Klickpfade, die Kategorien und die Bewertungen in relativ strukturierter Form vorliegen und somit eine automatisierte Analyse ohne große Datentransformation möglich ist.

3.3 Flickr-Clusters

Flickr bietet Nutzern die Möglichkeit, Bilder hochzuladen, die auch von anderen Nutzern betrachtet werden können. Zur Orientierung und Kategorisierung können Tags vergeben werden. Als problematisch erweist sich hierbei, dass diese Tags homonym oder synonym verwendet werden können. Zur Reduktion unerwünschter Suchergebnisse aufgrund von Homonymen wurden deshalb mit Methoden der Segmentierung Cluster eingeführt, sodass für jede Bedeutung eines Begriffes ein Cluster erstellt werden kann.[13] Beispielsweise werden zu dem Begriff „Turkey“ die Cluster bzw. Begriffsbedeutungen Türkei, Thanksgiving/Geflügelzubereitung und Vögel geführt.[14] Ein weiterführender Ansatz zu dieser Thematik findet sich bei LAWRENCE ET AL.; hier wird eine Methode zur Steigerung des Formalisierungsgrades von Tags vorgestellt.[15]

---

11 Vgl. Linden, G., Smith, B., York, J. (2003),S. 76-80.

12 Vgl. Agrawal, R., Imielinski, T., Swami, A. (1993), S. 207 f.

13 Vgl. Butterfield, S. (2005).

14 Vgl. http://flickr.com/photos/tags/turkey/clusters.

15 Vgl. Lawrence et al. (2007).

4 Potenziale des Einsatzes von Data Mining in User-Generated Content

4.1 Verwendungsmöglichkeiten aus ökonomischer Sicht

Tag Mining

Für Tags erscheinen die Einsatzpotenziale des Data Minings in User-Generated Content am größten, da diese einen vergleichsweise hohen Strukturierungsgrad aufweisen. Informationen, die in Tag Clouds oder in den Klickraten von Tags von Videoportalen, Internetseiten mit Produktbewertungen oder Social Bookmark-Sites enthalten sind, können beispielsweise zur Generierung von personalisierter Werbung genutzt werden. Auch Cross-Selling-Angebote können durch nutzergenerierte Tags in personalisierter Form hilfreich sein. So ist es möglich, dass Flickr als Teil von Yahoo! bei bestimmten Tags automatisiert Links zu Angeboten von Yahoo! setzt. Des Weiteren lässt sich die Homonymieproblematik von Tags auch auf die Suche nach Internetseiten ausweiten, die Lösung einer automatisierten Generierung von Clustern wird beispielsweise bei der Suchmaschine Vivisimo umgesetzt. Analog könnten auch Social Bookmarks mit nutzergenerierten Tags versehen werden, die in Cluster unterteilt werden.

Text Mining

Als problematischer erweist sich die Mustererkennung in unstrukturierten Texten. Potenziale hierfür bieten sich z. B. in Profilen von Social Networking-Sites sowie Kundenwünschen und -bewertungen in Foren. Eine Auswertung von Interessensschwerpunkten der Nutzer von sozialen Netzwerkseiten können wichtige Anhaltspunkte zur personalisierten Werbung liefern. Hierbei sind jedoch neben der Problematik der relativen Unstrukturiertheit der Daten auch Datenschutzbestimmungen sowie Vorschriften zum unerlaubten Wettbewerb zu beachten.

Verfahren des Clusterings können beispielsweise dabei helfen, Wünsche und Bewertungen von (potenziellen) Nutzern zu strukturieren, z. B. im Rahmen von Marktforschungsstudien vor einer Neueinführung eines Produktes. Eine solche Kategorisierung kann eine Auswertung dieser nutzergenerierten Inhalte erleichtern.

4.2 Verwendungsmöglichkeiten aus nicht-ökonomischer Sicht

Bisher ist es teilweise unklar, aus welchen Gründen sich Anwender entscheiden, eigene Inhalte anderen im Internet zur Verfügung zu stellen.[16] Aus Sicht der Nutzer könnte eine stärkere Entlohnung durch Werbeeinblendungen bzw. „Sponsored Links“ zusätzliche Anreize bieten, selbst erstellte Inhalte zu publizieren. Der AdSense-Dienst von Google unterstützt diesen Gedanken und bietet auch privaten Nutzern die Möglichkeit, ihre Inhalte zu analysieren und zum Inhalt der Seite passende Werbelinks einzublenden. Der Nutzer profitiert im Gegensatz zu herkömmlichen Werbeeinblendungen davon, dass er i. d. R. an den Klicks auf diese Links beteiligt wird und ist somit daran interessiert, dass die Links eine möglichst hohe Affinität zu dem Inhalt der Seite aufweisen. Verbesserte Methoden des Data Minings (hier insbesondere Clustering oder Assoziationsanalyse) können eine weitere Verbreitung und eine Effizienzsteigerung ermöglichen.

---

[16] Zu dieser Fragestellung vgl. Oded, N. (2007).

Data Mining als Dienst oder Commodity

Zu den Möglichkeiten, die als Dienst bzw. Commodity angebotene Data Mining-Verfahren bieten, zählen neben einer weiteren Diffusion von werbefinanzierten Entlohnungsmöglichkeiten und der Optimierung von Suchanfragen durch Tag Mining bei Social Bookmarks beispielsweise die Nutzung von Geokoordinaten und der damit zusammenhängenden Daten und die automatisierte Generierung von Meshups. Die Verwendung von Geokoordinaten und damit in Verbindung stehender Daten kann für Nutzer beispielsweise Vorteile bei einer Reiseplanung bieten. Die Webseite loc.alize.us[17] stellt einen Meshup aus Google Maps und Flickr zur Verfügung, in der Flickr-Fotos mit Geotags versehen werden und z. B. dem Reiseziel nahe gelegene Sehenswürdigkeiten identifiziert werden können. Des Weiteren können durch quelloffene Application Programming Interfaces Mash-Ups generiert werden, die sich automatisiert den Vorlieben eines Nutzers anpassen. Dies könnte durch Aufzeichnung seines Klickverhaltens, seine Bookmarks durch Tracking der im Internet gekauften Produkte geschehen.

5 Zusammenfassung und Ausblick

Die Zunahme der aktiven Teilnahme von Internetnutzern und der damit verbundenen Menge an nutzergenerierten Inhalten bietet vielfältige Möglichkeiten für den Einsatz von Data Mining-Verfahren. Zum Teil werden diese Inhalte bereits ökonomisch genutzt, um beispielsweise personalisierte Werbeeinblendungen zu präsentieren. Künftig sind jedoch neben allgemeinen Problemen des Data Minings, wie einem zur Analyse ungeeigneten Strukturierungsgrad der Daten, zusätzliche Fragestellungen zu erörtern, wie der Umgang mit Datenschutzbestimmungen oder unerwünschter Werbung.

Am Beispiel der Internetangebote von Amazon und Flickr wurden Praxisfälle vorgestellt, die verdeutlichen, dass Data Mining als nutzenstiftendes Verfahren für nutzergenerierte Inhalte auch wirtschaftlich sinnvoll eingesetzt wird. Eine Verbesserung und Verbreitung dieser Technologien könnte es auch privaten Nutzern ermöglichen, die Wahrnehmung ihres Internetangebotes zu erhöhen und somit auch finanziell zu profitieren.

Als zu lösende Problemstellungen sind auf der einen Seite Datenschutzbestimmungen zu beachten; ein Nutzer muss beispielsweise stets im Voraus der Auswertung seiner Daten zu Werbezwecken zustimmen. Zum anderen steht einer intensiven Analyse von User-Generated Content die hierbei häufig anzutreffende Unstrukturiertheit der Daten entgegen. Semantische Suchverfahren oder Methoden einer (teil)automatisierten Verschlagwortung bieten jedoch vielversprechende Möglichkeiten, diese Barrieren zu überwinden.

---

[17] Vgl. http://loc.alize.us.

Literaturverzeichnis

Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Databases, in: ACM SIGMOD Record, 22 (1993) 2, S. 207-216.

Beck, A.: Web 2.0: Konzepte, Technologie, Anwendungen, in: HMD – Praxis der Wirtschaftsinformatik, 255 (2007), S. 5-16.

Bensberg, F.: Web Log Mining als Instrument der Marketingforschung – Ein systemgestaltender Ansatz für internetbasierte Märkte, Wiesbaden 2001.

Butterfield, S.: The New New Things. 2005. http://blog.flickr.com/en/2005/08/01/the-new-new-things.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery: An Overview, in: Advances in Knowledge Discovery and Data Mining. Hrsg.: U. M. Fayyad et al. 1996, Menlo Park, Cambridge, London 1996, S. 1-33.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P.: Knowledge Discovery and Data Mining: Towards a Unifying Framework, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Hrsg.: E. Simoudis, J. Han, U. M. Fayyad, Menlo Park 1996, S. 82-88.

Fayyad, U. M.: Mining Databases: Towards Algorithms for Knowledge Discovery, in: Data Engineering, 21 (1998), 1, S. 39-48.

Hagedorn, J., Bissantz, N., Mertens, P.: Data Mining (Datenmustererkennung): Stand der Forschung und Entwicklung, in: Wirtschaftsinformatik, 39 (1997) 6, S. 601-612.

Lawrence et al.: Integrating Semantic Metadata Into an Existing Tagging Community: Semantic Tags Online. 2007. http://interaction.ecs.soton.ac.uk/faith/papers/tagging2.pdf.

Linden, G., Smith, B., York, J.: Amazon.com Recommendations: Item-to- Item collaborative filtering, in: IEEE Internet Computing, 7 2003 1, S. 76-80.

O'Reilly, T.: What is Web 2.0. 2005. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

Oded, N.: What motivates Wikipedians, or how to increase user-generated content contribution. Communications of the ACM, im Erscheinen (http://faculty.poly.edu/ %7Eonov/ Nov_Wikipedia_motivations).

OECD: Participative Web: User-Generated Content. 2007. http://www.oecd.org/dataoecd/ 57/14/38393115.pdf.

Rivadeneira et al.: Getting our head in the clouds: toward evaluation studies of tagclouds, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 07), San Jose 2007, S. 995-998.

Sifry, D.: The State of the Live Web, April 2007. 2007. http://www.sifry.com/alerts/archives/000493.html.

Szugat, M.; Gewehr, J.E.; Lochmann, C.: Social Software. Frankfurt a. M. 2006.

II Business-Aspekte

AdSense, verwandte Geschäftsmodelle und ihre Long-Tail-Effekte

Michael Räckers

1 Zielgerichtete Werbung und Geschäftstätigkeiten im Internet

Advertising, im Deutschen mit *Werben* oder auch *Anzeigen* zu übersetzen, beschreibt im Allgemeinen die Übertragung der klassischen Art der Werbung auf das Internet. Die bisherigen Medien der Wahl für die Anbieter waren Radio, Fernsehen oder Zeitschriften. Das sich hierbei ergebende Problem ist aber, dass sich der Inserierende nie sicher sein kann, dass er mit seiner Anzeige auch auf Interessenten trifft. Wenn ein Anbieter für eine Anzeige mit einer bestimmten Auflage eine bestimme Summe x zahlt, kann er sich in keiner Weise sicher sein, dass seine Anzeige von genügend Interessenten gelesen wird. Dies aber ist entscheidend, um einen positiven Nutzen, sprich zusätzliche Kaufabschlüsse, aus der Werbeaktivität zu erzielen. Auch die einfache Nutzung des Internets als Werbeplattform, z. B. durch schlichte Bannerwerbung, schafft für dieses Problem keine Abhilfe. Werbebanner auf Webseiten sind in diesem Verständnis lediglich digitalisierte Werbebotschaften, die unpersonalisiert bereitgestellt werden. Das Besondere bzw. Neue am Keyword-Advertising ist, dass die Werbung mit bestimmten Ziel- oder Suchwörtern verbunden wird, die es erlauben, ganz gezielt Personen, die nach diesen Stichworten suchen, zu bewerben [1].

Google bietet mit seinen Konzepten des AdSense [2] und der AdWords [3] Anbietern die Möglichkeit, ihre Werbebotschaften besser und effizienter zu platzieren, indem Möglichkeiten angeboten werden, die es erlauben, gezielt am Produkt interessierte, mögliche Kunden anzusprechen.

Neben dem reinen Bewerben von Produkten über das Internet über Keyword-Advertising haben sich verschiedene weitere Geschäftsmodelle, die das Internet als Verkaufsplattform nutzen, entwickelt. Hier sind beispielhaft *Broker* zu nennen, die eine Plattform für Geschäftsabschlüsse bieten, allgemein aber auch jede andere Form von Intermediären, die Kontakte zwischen Verkäufern und Käufern ermöglichen.

Im Folgenden werden zunächst als wohl bekannteste Vertreter des Keyword-Advertisings die beiden Konzepte AdSense und AdWords von Google vorgestellt [4]. Hierbei wird berücksichtigt und beschrieben, wie im Rahmen dieser Konzepte der so genannte Long-Tail-Effekt eintritt, der besagt, dass durch die Nutzung des Internets auch mit Nischenprodukten, die sich wenig verkaufen, hinreichend große Umsätze erzielt werden können, da die Kosten für Vorhaltung, Lagerung und Versand sehr gering ausfallen. Anschließend wird kurz auf weitere Beispiele für Advertising-Suchmaschinen eingegangen. Im Weiteren werden Vorteile, Nachteile, Chancen und Risiken der Angebote vorgestellt und anhand von Beispielen mögliche Gefahren und Missbräuche aufgezeigt. Im folgenden Abschnitt werden andere internetbasierte Geschäftsmodelle präsentiert, die es ermöglichen, mit dem Medium des Internets (Mehr-)Einnahmen zu erzielen, bevor im letzten Abschnitt ein kurzes Fazit gezogen wird.

2 Google AdSense

Google AdSense ist ein kostenloses Angebot, das es ermöglicht, mit eigenen Internetauftritten Geldeinnahmen zu erzielen [5]. Hierbei werden Werbebotschaften auf den eigenen Webseiten eingebunden, die zu den Inhalten und Angeboten der eigenen Seite passen. Für jeden Klick, der von Besuchern der eigenen Seiten auf den eingeblendeten Links durchgeführt wird, zahlt Google einen Beitrag aus. Dieses Verfahren nennt sich *Cost* oder auch *Pay-per-Click* [6], da nur im Falle eines Anklickens des eingeblendeten Links Beiträge fällig werden. Welche Werbelinks auf der eigenen Webseite eingeblendet werden, liegt nicht in der Hand des Webseitenbetreibers, sondern in der Hand von Google. Hierbei analysiert ein Crawler [4, 2] die Seite, auf der die Werbebotschaften (Ads) angezeigt werden sollen und berücksichtigt neben den reinen Inhalten der eigenen Seite auch Faktoren wie Herkunft und Sprache der Besucher. Für den Webseitenbetreiber gibt es aber Möglichkeiten und Einstellungen, die verhindern, dass Werbebotschaften z. B. von direkten Konkurrenten angeboten werden.

Soll die eigene Webseite durch Google AdSense Geldeinnahmen produzieren, gilt es zunächst zu ermitteln, ob die eigene Webseite den Programmrichtlinien von Google entspricht und in einer Sprache angeboten wird, die Google AdSense unterstützt [7]. Darüber hinaus gibt es einige Verhaltensregeln zu beachten [8]. Diese betreffen zunächst den Inhalt der eigenen Webseite. Die meisten dieser Einschränkungen sind dem Firmenstandort von Google – die USA – geschuldet, z. B., dass über die eigene Seite keine alkoholischen Getränke vertrieben werden dürfen.

Diese Programmrichtlinien besagen weiterhin, dass es grundsätzlich untersagt ist, auf der eigenen Webseite auf die Anzeigen zu klicken. Ebenso ist der Einsatz von Tools zur Generierung von Klicks verboten. Darüber hinaus schreiben die Richtlinien klar fest, dass keine anderen Personen zum Klick auf Anzeigen auf der eigenen Webseite aufgefordert werden dürfen. Kurzum, es sollen lediglich Klicks aus wirklichem Interesse an den verlinkten Produkten oder Inhalten entstehen. Ebenso dürfen keine expliziten Aufforderungen auf das Werbeangebot auf der Webseite erfolgen. Google überprüft das Klickverhalten auf den Webseiten dahin gehend sehr genau und sperrt schon beim Verdacht auf Missbrauch den entsprechenden Account [9]. Auf diese Problematik wird im späteren Teil dieses Artikels vertiefend eingegangen.

Wenn die Programmrichtlinien von der eigenen Webseite erfüllt werden, ist die Registrierung und Erzeugung eines Accounts für Google AdSense recht einfach, die Freischaltung der Seite erfolgt dennoch erst nach einer Prüfung durch Google [10]. Anschließend gibt es verschiedene Gestaltungsmöglichkeiten für die Einblendung der Anzeigen auf der eigenen Webseite. Die Größe des Werbefelds sowie die Position lassen sich ebenso konfigurieren wie die Farbgestaltung etc., sodass eine ergonomische Einbindung in den eigenen Auftritt möglich ist. Am Ende bietet AdSense ein JavaScript-Codefragment an, welches in den HTML-Code der eigenen Seite zu integrieren ist.

Zur Überwachung der Klicks auf Werbung auf der eigenen Webseite bietet Google verschiedene Reports an, die im eigenen Accountbereich abgerufen werden können und die Auskunft über Klicks und Einnahmen geben. Hierbei gilt es zu beachten, dass nicht jeder Klick auf eine Anzeige auf der eigenen Webseite gleich viele Geldeinnahmen bringt. Vielmehr hängt dies von den Zahlungen derjenigen ab, deren Webseiten über die eigene Webseite angesteuert werden.

Anzumerken an dieser Stelle ist noch, dass Google – Stand Januar 2006 – etwa 78,5 der Einnahmen an seine AdSense-Kunden auszahlt [11].

Wie aber bestimmt Google die Anzeigen, die auf den Webseiten geschaltet werden, und bekommt die Einnahmen von den angezeigten Anbietern? Die Anbieter müssen dazu das Konzept der *Google AdWords* nutzen und dort ihre Produkte verknüpfen, um in die Auswahl für das AdSense zu gelangen. Wie Google AdWords funktioniert, wird im nächsten Kapitel vorgestellt.

3 Google AdWords

Google AdWords ist ein Angebot, das es ermöglicht, auf Basis bestimmter Schlüsselwörter, die erworben werden können, Werbung auf Webseiten zu schalten. Das Prinzip von AdWords ist das eines Direktmarketings, da das Angebot zu möglichen Kunden nach dem Push-Prinzip geleitet wird. AdWords sind Anzeigen, die bei der Eingabe eines oder mehrerer Suchwörter (Suchwortgruppe) geschaltet werden (vgl. Abbildung 1, Links zu Tagungshotels aufgrund des Suchworts). Bei einer Google-Suchanfrage erscheinen diese in der Regel rechts neben den eigentlichen Suchergebnissen. Dieses Vorgehen ermöglicht es Anbietern, ihre Produktwerbung themenbezogen im Internet zu platzieren. Diese Anzeigen sind durch den Hinweis „Anzeigen“ deutlich von den anderen Suchergebnissen abgegrenzt.

Natürlich geschieht es, dass mehrere Anbieter oder Unternehmen sich das gleiche AdWord zur Schaltung einer Anzeige sichern möchten, dies ist möglich. Die Positionierung der Anzeigen wird hierbei über ein Auktionssystem entschieden [12]. Hierbei legt jeder, der ein Wort mit Werbeanzeigen hinterlegen möchte, ein Höchstgebot fest. Die Positionierung erfolgt aber nicht strikt nach der Reihenfolge der Gebote, sondern basierend auf Googles *PageRank-Algorithmus* [13] derart, dass Google für sich die größten Werbeeinnahmen erzielt. Über die *click-through-rates* und die Angaben z. B. zum maximalen Budget, welches man pro Monat bezahlen möchte, ermittelt Google die optimale Positionierung aus seiner Sicht. Dies kann bedeuten, dass ein Anbieter, der pro Klick weniger bezahlen möchte, aber ein höheres Budget pro Monat hat bzw. eine höhere click-through-rate, durchaus vor einem Anbieter gelistet werden kann, der pro Klick einen höheren Preis zahlt, wenn insgesamt mehr Einnahmen durch mehr erwartete Klicks beim weniger Bezahlenden erwartet werden können [14]. Hierzu sei ergänzt, dass keineswegs pro Klick das eigene Höchstgebot an Google entrichtet werden muss, sondern lediglich ein Cent mehr als das nächsthöhere Gebot. So kann es sein, dass für das Wort *Automobil* im Durchschnitt 1,02 € zu entrichten sind, für das Wort *Wirtschaftsinformatik* 1,10 € und für das Wort *ERCIS* lediglich 5 Cent, auch wenn man bereit ist, bis zu 80 € pro Klick zu bezahlen (Berechnung auf Basis von [15]). Weiterhin kann ein Tageshöchstsatz bestimmt werden, der – von Google durch Positionierung der Anzeigen passend optimiert – auf einen Monat betrachtet (voraussichtlich) nicht überschritten wird.

Neben der Kostenbeschränkung bietet sich als weitere Optimierung die Möglichkeit, das Erscheinen der eigenen Anzeige mit GoogleMaps bzw. über eine Entfernung vom eigenen Standort zu lokalisieren. Diese Faktoren gilt es im Besonderen bei der Bestimmung der eigenen Zahlungsbereitschaft zu berücksichtigen. Hier muss von realistischen prozentualen Kaufabschlüssen pro Klick ausgegangen werden, da bei einer entsprechend guten Positionierung in den Anzeigen viele Klicks auf die eigene Seite erzeugt werden, die aber noch längst nicht alle zu

Kaufabschlüssen führen. Beispielsweise sollten bei einem durchschnittlichen Kaufvolumen von 25 € und einer erwarteten Verkaufsrate von 1 % durch diese Klicks nicht mehr als 25 Cent pro Klick in der Auktion geboten werden.

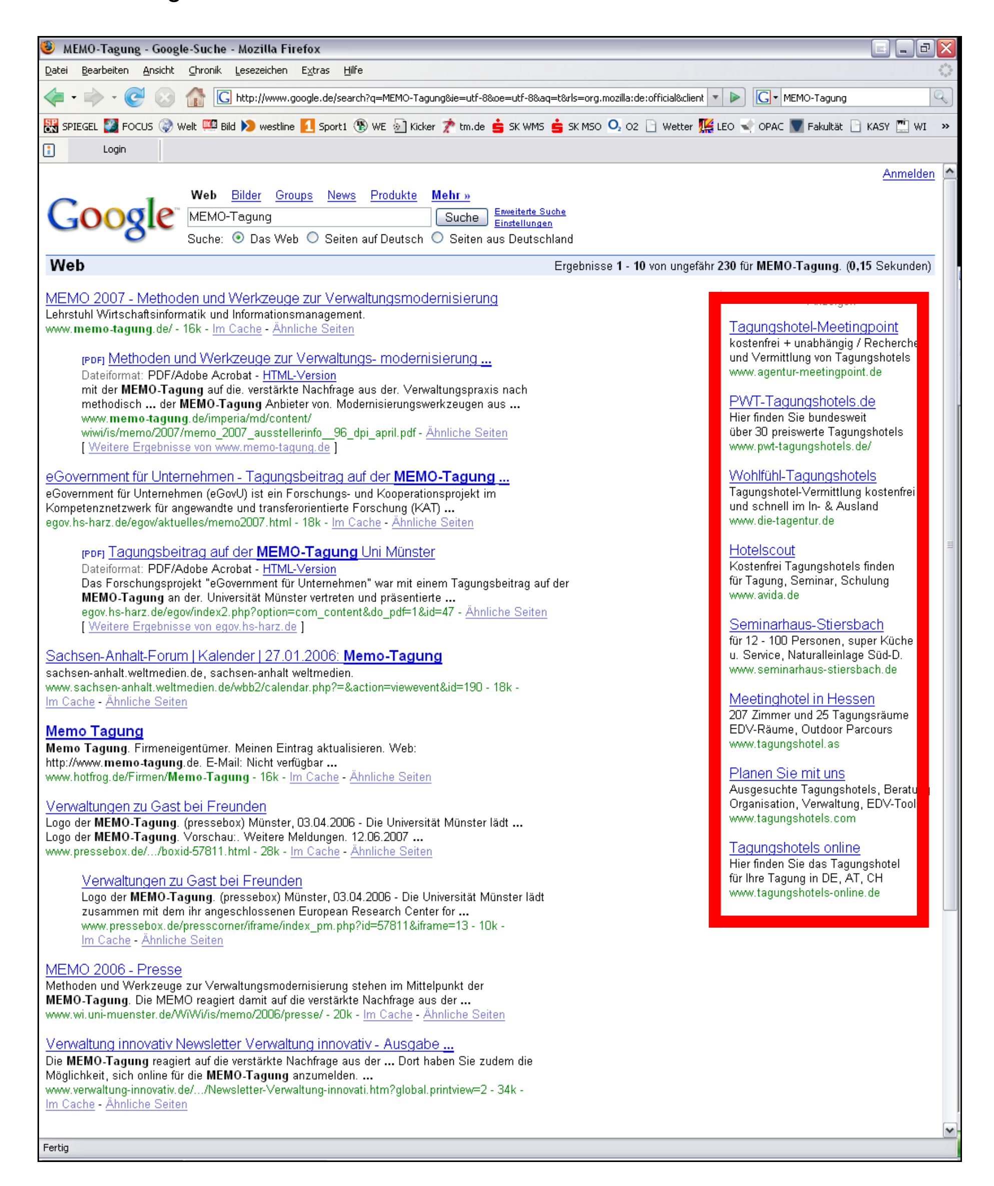


Abbildung 1: **Google AdWords-Angebote zum Suchwort „MEMO-Tagung“**

Was genau aber hat ein Anbieter davon, Produkte im Internet zu präsentieren und zu verkaufen? Die Frage stellt sich insbesondere bei Anbietern, bei denen nicht das komplette Geschäftsmodell auf dem Internet basiert, wie beispielsweise bei Amazon oder Dell [16, 17]. Chris Anderson umschrieb das Phänomen 2004 mit dem Begriff *Long-Tail-Effekt*, der besagt, dass ein Anbieter seinen Gewinn im Internet durch eine große Anzahl von Nischenprodukten erzielt [18, 19]. Dies funktioniert insbesondere aufgrund der so bezeichneten Demokratisierung der Produktionsmittel und der Vertriebswege durch das Internet. Breitbandtechnologie und PCs vereinfachen die Produktion und Lagerung z. B. von Musik und das Bereitstellen z. B. von sehr speziellen Büchern. So kann durch eine um ein Vielfaches größere Produktpalette der Gewinn gesteigert werden. Wenn ein Kunde bei iTunes [20] Musik kauft, ist egal, wo der Server mit der Datei steht. Es entstehen auch nicht wirklich Lagerkosten. So kann man nahezu kostenfrei ei-

nen viel größeren Kundenkreis erreichen als mit einem physischen bzw. herkömmlichen Geschäft, in dem CDs verkauft werden. Mit einer geschickten Marketingpolitik über Einträge in AdSense-Bereichen und hohe Listung bei den AdWords können so auch kleine Anbieter rund um die Welt Kunden erreichen und ihre Ware verkaufen. Wenn jemand also sein Geschäftsmodell nahezu vollständig auf den Vertrieb über das Internet ausrichtet, ist es nicht nur konsequent, sondern auch notwendig, das Marketing vollständig ins Internet zu verlagern. Auch und gerade in Verbindung mit immer mehr aufkommenden, themenbezogenen Blogs ergeben sich für Anbieter weit reichende ökonomische Möglichkeiten, sehr gezielt und effektiv Kunden anzusprechen. Da erfolgreiche Blogs von am Thema Interessierten häufig besucht werden, ist es geradezu ideal – wenn das Blog AdSense nutzt –, in den dortigen Anzeigen weit oben zu erscheinen. Die Chance, dass man auf wirklich interessierte Kunden trifft, ist sehr hoch. Durch geschicktes Bewerben sehr spezieller Suchwörter und damit verbunden geschicktes Leiten auf die betreffenden Unterseiten der eigenen Webseite wird ein Geschäftsabschluss pro Klick wahrscheinlicher. Darüber hinaus ist die Chance recht hoch, dass bei hoher Spezialisierung der Kreis der Konkurrenten in dieser Nische recht niedrig ist.

4 Weitere Advertising-Suchmaschinen-Anbieter

Es gibt eine Reihe weiterer Anbieter von Plattformen bzw. Suchmaschinen, die auf Basis des Pay-per-Click-Verfahrens arbeiten. Eine Auswahl von 13 Anbietern ist auf den Seiten des Pay-per-Click-Analysten [21] gelistet. Im Folgenden soll lediglich die „Urmutter“ dieser Art von Suchmaschinen vorgestellt werden, da sich diese vom Grundprinzip nicht unterscheiden.

GoTo.com

Als erste Advertising-Plattform gilt im Allgemeinen die Entwicklung „goto.com“ von Bill Gross in 1997. Mit GoTo wurde eine Plattform geschaffen, auf der Internetnutzer nach Dingen suchen konnten. Verbunden damit war das Geschäftsmodell des Pay-per-Click, was nichts anderes bedeutete, als dass diejenigen, die auf der Plattform gesucht werden konnten, Geld dafür bezahlten, dass über die GoTo-Seite auf ihre eigene Webseite geklickt wurde.

Die Reihenfolge der angezeigten Anbieter hing von Geboten ab, die für einen hohen Listenplatz abgegeben werden konnten. Auch das Konzept, dass Firmen oder Personen auf ihren Webseiten eine GoTo-Search-Box einstellen konnten und Geld für Suchen über diese eingebundenen Boxen bekamen, wurde zuerst auf dieser Seite angeboten. GoTo kann somit als absoluter Pionier in diesem Bereich des Advertisings gesehen werden. Erst zwei Jahre nach Beginn, ab 1999, kamen sukzessive andere Anbieter von solchen Seiten hinzu. GoTo nannte seinen Dienst im Jahr 2000 in Overture um und ging schließlich im Jahr 2003 in Yahoo auf [6]. An dieser Stelle sollte erwähnt werden, dass für eine Listung in der inzwischen unter Yahoo Search Marketing firmierenden Suchmaschine eine „Grundgebühr“ von 20 $ im Monat entrichtet werden muss. Im Vergleich: Bei Google AdWords sind dies lediglich 5 $.

Da Bill Gross sich sein Prinzip der Advertising-Suchmaschine im Juli 2001 hat patentieren lassen, kam es in der Folge im Jahr 2002 zu Klagen gegen FindWhat.com (inzwischen MIVA) und – was sicherlich mehr Aufsehen erregte – gegen Google, als diese AdWords auf den Markt brachten. Im Speziellen ging es bei der Klage gegen Google um den Auktionsmechanismus um

die besten Listenplätze in den Suchergebnissen. Der Rechtsstreit wurde schließlich gegen einen Austausch von 2,7 Mio. Aktien eingestellt [22].

5 Vorteile, Nachteile, Probleme und Gefahren des Advertisings

Zunächst sollen an dieser Stelle ganz kurz die Vor- und Nachteile des Advertisings im Allgemeinen zusammengefasst werden [5], ehe auf Probleme und Gefahren eingegangen wird.

Die Vorteile von Advertising im Internet liegen auf der Hand: Es ist recht einfach, seine Produkte vielen möglichen Kunden vorzustellen. Google AdWords und die verschiedenen weiteren Suchmaschinen bieten wunderbare Plattformen. Für Betreiber von Webseiten ist es durch das Konzept des AdSense sehr einfach, Einnahmen durch die Webseite zu erzielen. Google im Speziellen achtet zwar darauf, dass die Seiten, auf denen Werbung erscheinen soll, gewisse Inhalte, auch an Menge, vorweisen, um nicht reine Werbeplattformen zu erzeugen und wirklich inhaltsbezogen Werbung anzubieten. Ansonsten ist AdSense einfach bedienbar und erspart einem die aufwendige Suche nach Werbepartnern für die eigene Webseite. Gleichzeitig erweitert dies Konzept die Attraktivität von AdWords, da die Anzeigen nicht nur neben den Googletreffern erscheinen, sondern auch auf anderen Seiten. Dies erhöht natürlich die Zielgruppenerreichbarkeit.

Abseits.de [5] führt einige spezifische abrechnungstechnische und technische Nachteile von AdSense auf, insbesondere war aus dem Abrechnungssystem von Google z. B. lange nicht ersichtlich, von welchen Unterseiten des eigenen Webauftritts auf Werbelinks geklickt wurde. Bei der Verwaltung von mehreren Webauftritten aus einem Account heraus war es sogar anfangs nicht möglich, die Einnahmen den Webseiten zuzuordnen.

Ein besonderes Problem, welches mit der Einführung von über AdWords, Overture und anderen Suchmaschinen hinausgehenden AdSense-Konzepten sehr stark anstieg, ist der Versuch, die Werbenden zu betrügen. Es ließ sich beobachten, dass AdWords-Kunden, die vorher kleine monatliche Beträge für Klicks zahlen, durch die Einführung von AdSense erheblich mehr Klicks und Zahlungen hatten, allerdings kaum mehr Geschäftstätigkeiten verzeichnen konnten [23]. Der klassische Betrug wird durch so genannte *Click-Frauds* [24, 25] verübt. Hierbei werden Klicks auf Werbeeinblendungen durchgeführt, die in keiner Weise das Ziel haben, ein beworbenes Produkt zu erwerben, sondern nur darauf aus sind, demjenigen, der die Seite als Werbeplattform zur Verfügung stellt, die Einnahmen des Klicks zu bescheren. Schätzungen im Jahr 2005 zufolge sind etwa 20 % der Klicks betrügerisch [26].

Dies geschieht in der Regel nicht manuell, inzwischen ist die Entwicklung – auch aufgrund der Schutzmechanismen von Anbietern wie Google oder Yahoo – übergegangen zu ausgefeilten Click-Bot-Systemen [27]. Hierbei handelt es sich um Softwareprogramme, die Klicks vortäuschen und dabei – um Spuren so gut es geht zu verwischen – verschiedenste Absender-IP-Adressen und damit Herkunftsländer und -orte vortäuschen.

Als ein Beispiel für Click-Frauds, der durchaus wohl automatisiert erfolgte, kann der Erfahrungsbericht von Loren Baker dienen [28]. Er zeigt auf, wie sich die Menge der Klicks nach ei-

ner Registrierung bei Kanoodle.com deutlich erhöhte. Eine Analyse zeigt, dass die Besuchszeit der Klicks auf seiner Seite alle bei 0 Sekunden lagen, was natürlich höchst verdächtig ist.

Ein weiterer Erfahrungsbericht beschreibt den Betrug an einer Softwarefirma (2K Medical). Analysen der Logfiles ergaben ein wiederkehrendes Muster an Klicks auf die eigene Webseite auf Basis von nicht „normal" gesuchten Suchworten. In diesem Fall kamen fast alle IP-Adressen aus dem gleichen Land, in dem sich ein ehemaliger Mitarbeiter der Firma als Konkurrenzunternehmen selbstständig gemacht hatte. Der Schaden in diesem Fall lag bei 100.000 $ [29]. Dieses Beispiel zeigt im Besonderen, wozu Click-Frauds im Verhältnis von Firmenkonkurrenz dienen können bzw. genutzt werden. Da durch die erhöhte Menge an Klicks keine Verkäufe und damit keine zusätzlichen Einnahmen entstehen, sondern nur zusätzliche Ausgaben im Marketingbereich, soll ein Konkurrent auf dem Markt gezielt geschwächt werden. Genau wie im herkömmlichen Wettbewerb stellt sich an dieser Stelle die ethische Frage des Umgangs miteinander im Konkurrenzverhältnis, zumal es trotz Nachvollziehbarkeit von IP-Adressen etc. in keinem Fall leichter als im „Offlinebereich" nachzuweisen sein wird, dass diese Attacken gezielt von einem konkurrierenden Unternehmen lanciert wurden.

Betrügerische Vorgänge finden aber keineswegs nur mit dem Ziel statt, direkt einem bestimmten Unternehmen zu schaden, welches Werbung schaltet. Vielmehr versuchen Unternehmer oder Personen, die Werbung auf ihren Webseiten nach dem Cost-per-Click-Prinzip einstellen und für jeden Klick Geld bekommen, die Klicks zu ihren Gunsten in die Höhe zu treiben. Dies geschieht mit gleichen Mechanismen wie zuvor beschrieben. Ein Beispiel hierfür ist der ehemalige Google AdSense-Partner „Auction Experts International" aus Houston, Texas. Dieser erhielt 50.000 $ für Klicks, die über seine Seite auf Werbeeinblendungen durchgeführt wurden. Google verklagte das Unternehmen aufgrund von Unregelmäßigkeiten bzw. dem Verdacht des Klickbetrugs. Die Sachlage erschien wohl so eindeutig, dass Auction Experts International nie vor Gericht erschienen und Google somit die Klage gewann. Die Webseite des Unternehmens wurde aus dem AdSense-Partnerbereich gestrichen und ist nicht mehr existent [29].

In diesem Beispiel mag die Klage durchaus zu Recht erhoben worden sein und das Unternehmen zu Recht aus dem Partnerbereich entfernt worden sein. Insgesamt jedoch wird eine sehr große Intransparenz beklagt. Adam Penenberg griff 2005 in einem Artikel ein Zitat von Ben Edelmann auf, dass diese Situation gut zusammenfasst: „Search engines don't have the incentive but have the data, while the advertisers have the incentive but not the data." [29]

Überhaupt werden gerade die Geschäftsgebaren von Google in diesem Zusammenhang kritisch beäugt. So beklagt Microsoft im Zusammenhang mit der Übernahme von DoubleClick.com durch Google, dass ein Werbemonopol durch Google droht [30]. In diesem Fall mag das Wehklagen darauf zurückzuführen sein, dass Microsoft selbst an der Übernahme interessiert war, allgemein jedoch zeigt, dass 95 % aller Suchanfragen über Google abgewickelt werden, wie mächtig Google in diesem Bereich ist [31]. Um in diesem Zusammenhang noch einmal auf das Beispiel des ehemaligen AdSense-Partners „Auction Experts International" einzugehen bzw. ein weiteres Beispiel, PinkNews.co.uk [32], zu nennen: Was wäre, wenn die Klicks nicht betrügerisch herbeigeführt worden sind und Google trotzdem das Unternehmen aus seinem AdSense- bzw. AdWords-Bereich streicht? Hiawatha Bray berichtete 2003 über einen Fall, wo Google einen AdSense-Partner ohne Angabe von Gründen kündigte [33]. Vermutlich wurde diese Kündigung wegen des Verdachts des Klickbetrugs ausgesprochen, beweisen kann bzw. konnte Google dies wohl nicht.

Die Möglichkeiten, über das Internet keywortbezogen Werbung zu schalten, werden durch einen Ausschluss erheblich minimiert. Zwar gibt es weitere Anbieter, der Zielmarkt ist aber deutlich eingeschränkt, wie der Marktanteil von Google zeigt. Unter Berücksichtigung des möglichen Long-Tail-Effekts wird deutlich, dass dies gerade für echte Nischenproduzenten und -anbieter, deren Geschäftsmodell nahezu ausschließlich auf der großen Erreichbarkeit über das Internet basiert, ein Problem wäre. Dies kann ohne Weiteres existenzbedrohend sein.

6 Weitere internetbasierte Geschäftsmodelle

Neben den auf Advertising bezogenen Geschäftsmodellen von Google gibt es weitere Anbieter, die sich – vor oder nach Google – das Internet als Geschäftsplattform zueigen gemacht haben. Auch diese sollen an dieser Stelle kurz Erwähnung finden.

Über Werbeangebote hinaus bieten so genannte Broker die Möglichkeit für Verkäufer und Käufer, auf einer Internetplattform Geschäfte durchzuführen [12]. Das Geschäftsmodell des Brokers basiert auf Gebühren, die er für Angebote und/oder Transaktionen, die über seine Plattform abgewickelt werden, kassiert. Zu nennen sind hier beispielsweise Amazon Marketplace [16] oder auch eBay [34]. Bei Amazon zum Beispiel ist das Einstellen von Artikeln im Bereich des Marketplace kostenlos, bis der Artikel veräußert wurde. Das Geschäftsmodell von Anbietern, wie eBay, Alibaba [35] oder mySimon [36], basiert nur auf diesem Prinzip.

An dieser Stelle könnten unzählige weitere Beispiele für Broker genannt werden. Quasi für jede Art von Gütern gibt es auf spezielle Branchen bezogene Anbieter von Plattformen, die es erlauben, mit bestimmten (meist gebrauchten) Waren zu handeln. Man denke nur an die Automobilbranche [37, 38] oder an die Immobilienbranche [39-41] sowie Hotelbuchungssysteme [42, 43]. Alle diese Plattformen haben gemeinsam, dass sie (private) Verkäufer und (private) Nachfrager zusammenbringen.

Betrachtet man im Folgenden die Long-Tail-Effekte, die sich hier ergeben, so kann herausgestellt werden, dass sowohl Anbieter als auch Nachfrager von diesen Plattformen profitieren. Ganz klassisch im Sinne des Effekts erreichen die Anbieter mehr potenzielle Kunden, aber auch die Nachfrager haben mehr Möglichkeiten der Auswahl. Hier gilt es, für die Anbieter die größere Konkurrenz zu beachten. Ein Automobil als solches oder eine Immobilie als solche würde eben nicht als klassisches Nischenprodukt bezeichnet werden können, der Wettbewerb ist daher groß. Vielmehr ist hier eine Unterscheidung in sehr spezifische, spezielle Angebote ein Schlüssel. Nicht allgemeine Wörter wie „Immobilie“ gilt es zu bewerben, sondern komplexere Suchabfragen wie „Haus Garten „im Grünen“ ländlich“. Darüber hinaus sollte dann eine Weiterleitung bei Klick nicht auf die eigene Hauptseite verweisen, sondern am besten schon direkt auf die entsprechenden Angebote soweit möglich. So kann man sich gegenüber Konkurrenten möglicherweise differenzieren und erhöht die Chancen, dass der Besucher auf der Seite bleibt, da er nicht erst noch all seine Suchbegriffe erneut eingeben muss.

Das bereits erwähnte Amazon, aber auch buch.de [44] sind im Vergleich zu den genannten Plattformen ausschließlich oder darüber hinaus klassische Onlinestores, in denen selbst Produkte wie Bücher oder Elektronik angeboten werden. iTunes [20] verlangt weitergehend z. B. noch die Installation einer speziellen Software zum Erwerb der angebotenen Musik. Im Gegen-

satz zum Broker wird hier nicht lediglich die Plattform für die Bereitstellung und Abwicklung zur Verfügung gestellt. Als weitere Beispiele können hier Schlecker [45], dm-Drogeriemarkt [46] oder Conrad Electronic [47] stellvertretend für viele Weitere genannt werden.

Für diese Onlinestores liegt der Long-Tail-Effekt insbesondere in der Erweiterung des Kundenstamms. Hier geht es nicht darum, besondere Nischen zu bedienen, sondern die Möglichkeit, den Umsatz als solchen zu steigern.

Als Extrem an dieser Stelle können die Unternehmen angeführt werden, die ihre Produkte ausschließlich über das Internet verkaufen. Sehr bekanntes Beispiel an dieser Stelle ist das Unternehmen Dell [17]. Dadurch, dass die angebotene Ware ausschließlich über das Internet angeboten wird, erspart sich Dell den Aufbau eines Filialnetzes und kann so Intermediäre wie Händler bzw. Einzelhändler komplett ausschalten. Der Gewinn kann zusätzlich abgeschöpft werden, sodass an dieser Stelle nicht lediglich ein Long-Tail-Effekt aufgrund eines zusätzlichen Verkaufskanals auftritt, sondern der Effekt für die kompletten Verkäufe gilt.

7 Zusammenfassung und Fazit

Advertising bzw. das Werben im Internet bietet eine Vielzahl an Möglichkeiten, um sich und sein Produkt zu platzieren. Insbesondere unter dem Aspekt des Keyword-Advertisings gelingt es zunehmend, ganz gezielt lediglich die am Produkt interessierten Internetnutzer auf seine Seiten zu lotsen. Die technischen Weiterentwicklungen im Bereich des Internets bieten eine Vielzahl von Möglichkeiten, Werbung zu platzieren. Dies führt zu einer effektiveren Nutzung der Werbe- und Marketingaufwände. Der Schwerpunkt der Ausarbeitung zeigt, dass Google an dieser Stelle – zumindest was die Marktpräsenz angeht – das Maß aller Dinge ist. AdSense bietet Betreibern von Webseiten und Blogs gute Möglichkeiten, mit den eigenen Seiten ökonomisch zu arbeiten und Geld zu verdienen.

Wo Geldeinnahmen winken, ist der Missbrauch allerdings nicht weit. Insbesondere bei den Pay- bzw. Cost-per-Click-Verfahren ist es verlockend, durch zusätzliche Klicks die Einnahmen in die Höhe zu treiben. Auch das bewusste Schädigen von Konkurrenten tritt in diesem Zusammenhang auf, welches unter sozialen Aspekten sehr kritisch zu betrachten ist. Hier muss ebenfalls die exponierte Stellung von Google kritisch hinterfragt werden. Aus Sicht der AdWords-Kunden ist ein strenges Durchgreifen beim Verdacht auf Klickbetrug zwar wünschenswert, die Gefahr des unbegründeten Ausschlusses von AdSense ist aber schnell gegeben. Da sich hier für Webseitenbetreiber allerdings ein Geschäftsmodell entwickelt hat, auf dessen Einnahmebasis sich finanziert wird, kann ein Ausschluss und Abschneiden der Einnahmequellen existenzbedrohend sein.

Dies gilt natürlich und im besonderen Maße auch für AdWords-Kunden, denen – z. B. von Konkurrenten – Klickbetrug vorgeworfen wird und die aufgrund dessen nicht mehr über AdWords werben dürfen. Insbesondere für rein auf dem Internet basierende Geschäftsmodelle ist das Werben im Internet aber existenziell! Das Internet hat sich zu einem neuen, virtuellen Marktplatz oder auch Flohmarkt entwickelt, der für viele kleine und exotische Anbieter die einzige Chance ist, ihr Geschäft zu betreiben. Hier spielt auch der Aspekt der Globalisierung eine große Rolle. Wie sonst wäre es einfach möglich, in Deutschland Produkte z. B. aus Asien zu bezie-

hen? Gerade in den Nicht-Metropolen sind entsprechende Geschäfte selten bis nicht vorhanden. In diesen Fällen kann Google durch seine Marktmacht eine sehr entscheidende Rolle spielen. Diese Rolle gilt es, aufgrund der vielfach kritisierten Intransparenz über die tatsächlichen und abgerechneten Klicks und der Kriterien für Ausschlüsse durch Google zu hinterfragen.

Literaturverzeichnis

[1] Wikipedia. (2007) Keyword Advertising – Wikipedia.[Online]. Available: http://de.wikipedia.org/wiki/Keyword_advertising.

[2] Wikipedia. (2007) Google AdSense – Wikipedia.[Online]. Available: http://de.wikipedia.org/wiki/AdSense.

[3] Wikipedia. (2007) Google AdWords – Wikipedia.[Online]. Available: http://de.wikipedia.org/wiki/AdWords.

[4] Milstein, S., Dornfest, R.: Google – The Missing Manual. 2nd edn. O'Reilly, Sebastopol, CA (2004).

[5] Abseits.de. (2007) Adsense. Vor- und Nachteile von AdSense. Eine Einführung von Abseits.de.[Online]. Available: http://www.abseits.de/adsense.

[6] Mordkovich, B., Mordkovich, E.: Pay-per-click Search Engine Marketing Handbook: Low Cost Strategies to Attracting New Customers Using Google, Yahoo & Other Search Engines. MordComm Inc., New York (2005).

[7] Google. (2007) Welche Sprachen werden von AdSense unterstützt?[Online]. Available: https://www.google.com/adsense/support/bin/answer.py?answer=9727&hl=de.

[8] Google. (2007) Google AdSense-Programmrichtlinien.[Online]. Available: https://www.google.com/adsense/support/bin/answer.py?answer=48182&sourceid=aso&subid=ww-ww-et-asui&medium=link.

[9] selfhtml.de. (2007) Sieben Todsünden für Google Adsense-Teilnehmer.[Online]. Available: http://www.selfhtml.de/artikel_1_1_sieben-tods %FCnden-google-adsense-ausschluss-teilnahme.htm.

[10] Miller, M.: Googlepedia: The Ultimate Google Resource. 1st edn. Que Publishing, Indianapolis, IN (2007).

[11] NewYorkTimes. (2007) Google's Shadow Payroll Is Not Such a Secret Anymore.[Online]. Available: http://www.nytimes.com/2006/01/16/technology/16ecom.html?ex=1295067600&en=762ce1a27484c0b1&ei=5090&partner=rssuserland&emc=rss.

[12] Vossen, G., Hagemann, S.: Unleashing Web 2.0 – from concepts to creativity. 1st edn. Elsevier, Burlington, MA (2007).

[13] Wikipedia. (2007) PageRank – Wikipedia.[Online]. Available: http://de.wikipedia.org/wiki/PageRank.

[14] Battelle, J.: Die Suche – Geschäftsleben und Kultur im Banne von Google & Co. Börsenmedien, Kulmbach (2006).

[15] Google. (2007) Google AdWords: Keyword-Tool.[Online]. Available: https://adwords.google.de/select/KeywordToolExternal?defaultView=1.

[16] Amazon. (2007) Amazon.de – Jetzt verkaufen![Online]. Available: http://Amazon.de/gp/seller/sell-your-stuff.html/ref=sv_w_5/302-3110115-7851221.

[17] Dell. (2007) Bei Dell nach Notebooks, Desktops, Servern, Druckern, Software, Monitoren und Services suchen.[Online]. Available: http://www1.euro.dell.com/content/default.aspx?c=de&l=de&s=gen.

[18] Anderson, C. (2007) Wired 12:10: The Long Tail.[Online]. Available: http://www.wired.com/wired/archive/12.10/tail.html.

[19] Wikipedia. (2007) Long Tail – Wikipedia.[Online]. Available: http://de.wikipedia.org/wiki/Long_Tail.

[20] Apple. (2007) iTunes – Laden Sie iTunes.[Online]. Available: http://www.apple.com/de/itunes/download/.

[21] PayPerClickAnalyst. (2007) Pay Per Click Analyst.[Online]. Available: www.payperclickanalyst.com.

[22] Olsen, S. (2007) Google, Yahoo bury the legal hatchet | CNET News.com.[Online]. Available: http://netscape.com.com/Google,+Yahoo!+bury+the+legal+hatchet/2100-1024_3-5302421.html.

[23] CMSRanking. (2007) Was ist Adsense Marketing?[Online]. Available: http://www.cms-ranking.de/adsense_marketing.html.

[24] Davis, H.: Google Advertising Tools – Cashing in with Adsense, Adwords and the Google APIs. 1st edn. O'Reilly, Sebastopol, CA (2006).

[25] Strocchiola, J. (2007) SearchEngineWatch: Lost Per Click: Search Advertising & Click Fraud.[Online]. Available: http://searchenginewatch.com/showPage.html?page=3387581.

[26] Krüger, A. (2007) IT News World – Google & Co. wegen falscher Klicks vor dem Kadi.[Online]. Available: http://www.it-news-world.de/news_587/Google+&+Co.+wegen+falscher+Klicks+vor+dem+Kadi.

[27] Helm, B. (2007) Click Frauds get Smarter.[Online]. Available: http://www.businessweek.com/technology/content/feb2006/tc20060227_930506.htm?campaign_id=search.

[28] Baker, L. (2007) Kanoodle and Click Fraud.[Online]. Available: http://www.searchenginejournal.com/kanoodle-and-click-fraud/2235/.

[29] Penenberg, A. (2007) Click Fraud: Problem and Paranoia.[Online]. Available: http://www.wired.com/culture/lifestyle/news/2005/03/66845.

[30] Krüger, A. (2007) IT-News World – Microsoft beklagt drohendes Google-Werbemonopol.[Online]. Available: http://www.it-news-world.de/news_993/Microsoft+beklagt+drohendes+Google-Werbemonopol.

[31] Krüger, A. (2007) IT-News World – Google bleibt Spitzenreiter.[Online]. Available: http://www.it-news-world.de/news_1010/Google+bleibt+Spitzenreiter.

[32] TimesOnline. (2007) The nonsense about AdSense – Times Online.[Online]. Available: http://business.timesonline.co.uk/tol/business/markets/united_states/article703023.ece.

[33] Bray, H. (2007) Boston.com/Business/Technology/As Google grows, critics emerge.[Online]. Available: http://www.boston.com/business/technology/articles/2003/11/24/as_google_grows_critics_emerge/.

[34] eBay. (2007) eBay: Neue und gebrauchte Elektronikartikel, Autos, Kleidung, Sammlerstücke, Sportartikel und mehr – alles zu günstigen Preisen.[Online]. Available: http://www.eBay.de/.

[35] Alibaba. (2007) Alibaba Manufacturer Directory – Suppliers, Manufacturers, Exporters & Importers.[Online]. Available: http://www.alibaba.com/.

[36] mySimon. (2007) mySimon – Price Comparison Shopping.[Online]. Available: http://www.mysimon.com/.

[37] AutScout24. (2007) AutoScout24 Europas Automarkt für Gebrauchtwagen und Neuwagen.[Online]. Available: http://www.autoscout24.de/.

[38] mobile.de. (2007) mobile.de – der Automarkt für Gebrauchtwagen und Neufahrzeuge.[Online]. Available: http://www.mobile.de/.

[39] immobilienscout24. (2007) Wohnungen und Immobilien online – Immobilien Scout24.[Online]. Available: http://www.immobilienscout24.de/de/finden/wohnen/index.jsp;jsessionid=B24C78827A0B7DE42E0288C933E3A0CD.worker2.

[40] immonet.de. (2007) Immobilien Wohnungen Immobilien Wohnungen – Immonet.de.[Online]. Available: http://www.immonet.de/.

[41] immowelt.de. (2007) Immobilien Wohnungen Häuser – Immobilie, Wohnungen kaufen oder Mieten.[Online]. Available: http://www.immowelt.de/.

[42] hotel.de. (2007) hotel.de: 210.000 Hotels weltweit + Für Business & Privat + Internet-Bestpreise + Kostenlose Buchung.[Online]. Available: http://www.hotel.de/.

[43] HotelReservationService. (2007) HRS (Hotel Reservation Service.[Online]. Available: http://www.hrs.de/.

[44] buch.de. (2007) Internet-Shop für Bücher, Musik, Filme, E-Books, Hörbücher, CDs, DVDs und Software.[Online]. Available: http://www.buch.de/shop/home/show/.

[45] Schlecker. (2007) Schlecker.com.[Online]. Available: http://www1.schlecker.com/htdocs/Home.aspx?WasHere=TRUE&land=de.

[46] dm-Drogeriemarkt. (2007) dm-Homepage Deutschland.[Online]. Available: http://www.dm-drogeriemarkt.de/.

[47] ConradElectronic. (2007) Europas führendes Versandhandelsunternehmen für Elektronik und Technik.[Online]. Available: http://www.conrad.de.

SaaS-Geschäftsmodelle im Web 2.0

Sebastian Hallek

1 Revolution im Softwaremarkt durch Software as a Service?

Das Thema Software as a Service (SaaS) wird in der Praxis vielseitig und teilweise kontrovers diskutiert. In zahlreichen Marktforschungen wird das SaaS-Modell als eine interessante Alternative zum klassischen Lizenzmodell betrachtet, welche sich mittel- bis langfristig auf dem Markt etablieren werde. Einige IT-Analysten glauben, dass Software as a Service ein Milliardenmarkt wird [1], Marktanalysten der International Data Corporation sehen SaaS als eines der IT-Trendthemen für 2007 an [2] und GARTNER erwartet für SaaS im Unternehmensumfeld ein erhebliches Wachstum im Jahr 2007 [3] und prognostiziert, dass bis 2011 insgesamt 25 % der gesamten Unternehmenssoftware auf Basis des SaaS-Modells entwickelt werden [4].

Im Gegensatz dazu finden sich aber auch zahlreiche Meinungen, die sich kritischer mit dem SaaS-Modell auseinandersetzen und dieses nicht als neuen Trend, sondern – im Kontext des gescheiterten Application Service Providings (ASP) [VH07] – SaaS als die Wiedergeburt von ASP [5], als Wiederbelebung von ASP unter einem neuen Namen [6] oder auch als alten Wein in neuen Schläuchen [vGu07] sehen.

Aufgrund dessen wird im Folgenden mit der Analyse und Definition des SaaS-Modells sowie einer Einordnung im Kontext des ASP-Modells eine Grundlage für eine weitergehende Betrachtung des zugrunde liegenden Geschäftsmodells geschaffen, um dieses anschließend kritisch zu würdigen und daraus die Erfolgschancen für SaaS abzuleiten.

2 Software as a Service

2.1 Definition von Software as a Service

Die Begrifflichkeiten rund um SaaS werden in der Regel nicht eindeutig verwendet respektive voneinander abgegrenzt. Einige Autoren betrachten die Begriffe Software as a Service, Application Service Providing und On-demand-Services als Synonyme [6]. Darüber hinaus werden bei zahlreichen Diskussionen nur einzelne Aspekte akzentuiert, ohne eine umfassende Definition abzugeben [7].

Zwecks Ableitung einer geeigneten Definition von Software as a Service werden die beiden konstituierenden Begriffe Software und Service im Detail betrachtet.

Software bezeichnet dem allgemeinen Verständnis nach alle nicht-physischen Funktionsbestandteile eines Anwendungssystems [8].

Als Dienstleistung (Service) wird eine Leistung bezeichnet, die nicht der Produktion eines materiellen Gutes dient und bei der auch nicht der materielle Wert eines Endproduktes im Vorder-

grund steht [9]. Sie ist nicht lagerbar, nicht übertragbar und benötigt den externen Faktor Kunde.

Darauf aufbauend kann Software as a Service als ein Modell verstanden werden, nach dem eine Software als Service angeboten wird und dessen Nutzung somit dem uno-actu-Prinzip folgt: Produktion und Konsumption der Leistung fallen zusammen.

2.2 Konstituierende Merkmale und Implikationen

Eine weitergehende Betrachtung der oben genannten Definition von SaaS nach einigen konstituierenden Merkmalen verdeutlicht die impliziten Prämissen [10].

2.2.1 Hosting

Die Software wird von dem SaaS-Anbieter oder seinem Dienstleister betreut und gewartet. Der Nutzer wird nicht dazu aufgefordert, die Software selbst zu installieren oder zu betreiben, sondern bleibt von den üblichen Wartungsaufgaben befreit.

2.2.2 Publikation

Die zugrunde liegende Definition schreibt nicht die Form der Verbreitung der Software vor. Grundsätzlich wird der SaaS-Anbieter jedoch seine Software über das Internet zur Verfügung stellen und die gängigen Vorteile, wie das Aufsetzen auf bereits etablierte Standards oder die hohe Verbreitung des Internets, nutzen.

2.2.3 Mandantenfähigkeit

Der Definition folgend besteht die Idee von SaaS darin, die Software mehreren Kunden anzubieten und ggf. auch auf die Bedürfnisse des Kunden hin anzupassen.

2.2.4 Veredelung

Ein SaaS-Anbieter kann sich bei der Erbringung seiner Leistung auf andere, ihm angebotene Software beziehen und diese zu einer höherwertigeren kombinieren. Er muss die Fertigungstiefe – analog zur industriellen Produktion – nicht komplett beherrschen.

2.2.5 Entgelt

Bei der Inanspruchnahme der Software findet nicht das klassische Kaufmodell Anwendung, welches den einmaligen Erwerb von Nutzungsrechten vorsieht. Vielmehr fließt der tatsächliche Nutzungsgrad bei der Kalkulation ein.

2.3 Einordnung von Software as a Service, Application Service Providing und webbasiertes Application Service Providing

Bereits erwähnt wurde der Umstand, dass die Begrifflichkeiten im Umfeld von SaaS in der Literatur nicht eindeutig definiert werden. Als die zentralen Unterschiede werden in diesem Zusammenhang häufig die Möglichkeit der Abbildung von mehreren Kunden innerhalb der Software, d. h. die Mehrmandantenfähigkeit [vGu07], und die Bereitstellung resp. Veröffentlichung der Software über das Internet auf Basis von Webtechnologien genannt.

Im Gegensatz dazu stellt der zugrunde liegenden Definition von SaaS folgend ASP eine konkrete Form resp. eine Spezialisierung von SaaS dar. Die Idee von ASP ist ebenfalls die Bereitstellung einer Software über das Internet, allerdings wurden mitunter andere Paradigmen verfolgt und unterschiedliche Technologien verwendet.

Um den in der Literatur beschriebenen Unterschieden zwischen SaaS und ASP Rechnung zu tragen, werden die Merkmale von ASP im Vergleich zu denen des im Allgemeinen als SaaS bezeichneten Modells der internetbasierten Bereitstellung der Software herausgestellt. Letzteres wird zwecks Abgrenzung zur SaaS-Definition im Rahmen dieser Arbeit als Bereitstellung von Software als Service und unter Berücksichtigung der Webfähigkeit der Software als webbasiertes ASP (WASP) bezeichnet.

Kriterium	SaaS	
	ASP	WASP
Zielsetzung	Sukzessives Outsourcing von IT-Leistungen und Prozessen	Reines Outsourcing von IT-Leistungen
Komplexität der Software/ Einrichtung	Komplexe, spezifisch angepasste Software	Standardisierte Software mit geringen Anpassungen
Abbildung von Kunden	Nutzung kundenspezifischer Software (1:1-Beziehung)	Nutzung einer Software für alle Kunden (1:n-Beziehung)
Entgelt	i.d.R. variables Entgelt (nach Nutzungsvolumen)	i.d.R. variables Entgelt (nach Nutzungsvolumen)
Wartung und Betrieb	Wartung und Betrieb erfolgen durch den Anbieter	Wartung und Betrieb erfolgen durch den Anbieter
Softwarezugriff	Spezielle Zugriffsarchitektur, bspw. Terminalserver	webbasierter Zugriff per Browser

Abbildung 1: Charakteristika von SaaS-Modellen

Die in Abbildung 1 dargestellten Eigenschaften von ASP und WASP verdeutlichen, dass beide konkrete Ausprägungen der SaaS-Idee darstellen und sich nur in wenigen Punkten unterscheiden.

2.4 Erfolgschancen von Software as a Service

Im Allgemeinen werden die Chancen als sehr hoch angesehen, dass sich das Modell Software as a Service in naher Zukunft erfolgreich am Markt durchsetzen kann.

Mit dieser These geht einher, dass Software einem ähnlichen Preisverfall unterliegen wird wie Hardware in den letzten 40 Jahren [VH07]. Igniter Ltd. hat in diesem Zusammenhang eine Studie veröffentlicht, derzufolge ein Unternehmen für einen Arbeitsplatz ca. 33 € pro Monat aus-

geben müsste, um auf Basis von Open Source-Software oder von im Internet verfügbaren Services die notwendigen Funktionalitäten bereitzustellen.

Auf der anderen Seite ist unter Berücksichtigung der geringen Marktdurchdringung von ASP und der Ähnlichkeit zwischen ASP und WASP jedoch kritisch zu hinterfragen, ob der identifizierte Trend beibehalten wird und sich WASP am Markt durchsetzen kann. Es erscheint als zweckmäßig, diese Fragestellungen im Kontext des zugrunde liegenden Geschäftsmodells zu prüfen.

Zudem sind der zeitliche Kontext und die technischen Rahmenbedingungen, wie beispielsweise Internetnutzung und Brandbreite, zu beachten und auch betriebswirtschaftliche und soziale Aspekte, beispielsweise Zielsetzung, Verlustängste, Widerstand oder Sicherheitsaspekte, mit in eine detaillierte Evaluation aufzunehmen.

3 Internetbasierte Geschäftsmodelle

3.1 Definition eines internetbasierten Geschäftsmodells

Im allgemeinen Sprachgebrauch hat sich der Begriff Geschäftsmodell zur Beschreibung von unternehmerischen Tätigkeiten etabliert, welche auf die Finanzierung resp. Erlösgenerierung fokussiert.

Eine Dekomposition des Begriffs Geschäftsmodell legt eine spezifischere Bedeutung nahe. Geschäftsmodelle stellen demnach eine Aggregation wesentlicher relevanter Aspekte aus den betriebswirtschaftlichen Teildisziplinen dar, um einen einfachen, komprimierten Überblick über die Geschäftsaktivitäten in Modellform zu geben [Wir01]. Sie werden in Form mehrerer Partialmodelle beschrieben.

Die Besonderheit eines internetbasierten Geschäftsmodells besteht darin, dass die Marktleistung vollständig oder überwiegend mit Internettechnologien realisiert wird.

3.2 Komponenten von internetbasierten Geschäftsmodellen

In der Literatur sind mehrere Ansätze vertreten, die die Bestandteile eines internetbasierten Geschäftsmodells herleiten [AZ01, Stä01]. Diesen Ansätzen zufolge umfasst ein Geschäftsmodell eine Beschreibung der Unternehmensstruktur, Marktleistung, Leistungserstellung und des Erlösmodells. Zudem werden weitere Perspektiven von den unterschiedlichen Autoren akzentuiert.

Als äußerst zweckmäßig hat sich der Ansatz von WIRTZ herausgestellt, welcher eine Aufteilung eines Geschäftsmodells in die sechs Partialmodelle Marktmodell, Beschaffungsmodell, Leistungserstellungsmodell, Leistungsangebotsmodell, Distributionsmodell und Kapitalmodell – wie in Abbildung 2 dargestellt – vorschlägt [Wir01].

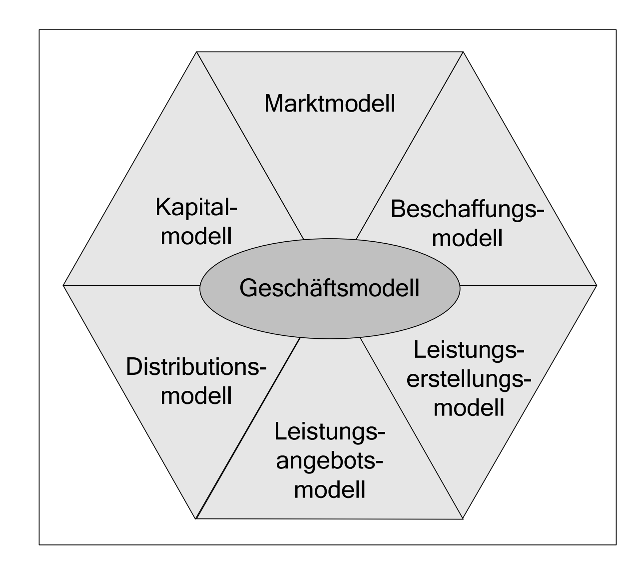


Abbildung 2: Partialmodelle eines integrierten Geschäftsmodells [Wir01]

Das *Marktmodell* beschreibt die Strukturen der relevanten Märkte und die bestehenden Marktteilnehmer. In diesem Umfeld können beispielsweise Wettbewerber, Nachfrager, Marktbedürfnisse und darauf aufbauend Marktsegmente betrachtet werden.

Das *Beschaffungsmodell* bildet das Verhältnis zu den Lieferanten ab.

Im *Leistungserstellungsmodell* werden die Prozesse betrachtet, mit denen die zur Verfügung stehenden Ressourcen, beispielsweise Dienstleistungen, zu der am Markt angebotenen Leistung transformiert werden.

Das *Leistungsangebotsmodell* beschäftigt sich mit der Fragestellung, welches Leistungsspektrum welchen Nachfragern bzw. Kunden angeboten wird.

Im *Distributionsmodell* wird festgehalten, welche Produkte und welche Dienstleistungen in welcher Weise zu welchem Zeitpunkt vom Anbieter zum Nachfrager transportiert werden.

Das *Kapitalmodell* zeigt aufgeteilt in ein Finanzierungs- und ein Erlösmodell, mit welchen Quellen das unternehmerische Tun und Handeln finanziert und welche Erlösformen genutzt werden sollen.

3.3 Beschreibung der WASP-Geschäftsmodelle

Im Folgenden werden potenzielle und am Markt auffindbare WASP-Geschäftsmodelle beschrieben.

Marktmodell

Eine Einschränkung des zu betrachtenden Marktes für WASP-Anbieter resp. die relevanten Marktteilnehmer ist nicht möglich. WASP stellt eine technologiegetriebene Entwicklung dar. Prinzipiell lässt sich jegliche Form von Anwendungssoftware über das WASP-Modell vertreiben und auch jegliche Kundengruppe ansprechen.

Hierbei stehen WASP-Anbieter vor der Herausforderung, ein geeignetes Marktsegment zu finden bzw. zu definieren, in dem eine hinreichend große Menge von Interessenten angesprochen, eine geeignete Finanzierung des Angebots sichergestellt und ein Wettbewerbsvorteil gegenüber anderen Anbietern klar kommuniziert werden kann.

Hierzu lassen sich mehrere interessante Beispiele für unterschiedlichste Segmente im Internet finden. Der Kalender *30Boxes* [11] oder die Tabellenkalkulation *NumSum* [12] stellen branchenunabhängige Produkte dar, welche für Privatpersonen und Unternehmen von Interesse sein können. *salesforce.com* ist im Gegensatz dazu ein Anbieter, der sich mit der gleichnamigen Software für das Customer Relationship Management auf das Business-to-Business-Segment konzentriert und branchenunabhängige Software vertreibt [13]. Eine brachenspezifische Software für den Onlineversandhandel stellt das Shopsystem *ON Demand GRID* der Firma *commercetools* dar [14].

Beschaffungsmodell

Sieht man von der Möglichkeit zur Komposition von mehreren Services zu einem höherwertigem Service ab, sind im WASP-Umfeld keine Besonderheiten hinsichtlich des Beschaffungsmodells erkennbar.

Prinzipiell befinden sich WASP-Anbieter in derselben Ausgangssituation wie die Anbieter klassischer Software, dass weniger die Zulieferungen anderer Unternehmen, sondern mehr das Fach- und IT-Know-how der Mitarbeiter über die Qualität der eigenen Wertschöpfung entscheidet.

Leistungserstellungsmodell

Die Leistungserstellung im WASP-Umfeld kann analog zum Modell des Wertschöpfungsprozesses für Dienstleistungen in die drei Phasen Erstellung der Leistungsbereitschaft, Leistungsvereinbarung und Leistungserbringung differenziert werden.

Die *Erstellung der Leistungsbereitschaft* umfasst im Wesentlichen die Entwicklung und Bereitstellung einer Software auf Basis von Webtechnologien, die von mehreren Usern genutzt und über das Internet angesprochen werden kann. Im engeren Sinn stellt dies den Kern der Wertschöpfung dar, da die Software ohne aufwendige Anpassungen durch Kunden genutzt werden soll [15]. In wenigen Fällen können anbieter- und/oder kundenseitige Anpassungen an der Software – wie beispielsweise im Fall individueller Felder bei *Salesforce.com* – erforderlich sein.

Im Rahmen der Erstellung der Leistungsbereitschaft sind darüber hinaus auch die Betriebs- und Wartungsarbeiten sowie die Weiterentwicklung der Software zu berücksichtigen. Letzteres stellt eine Besonderheit im WASP-Umfeld dar, da diese unter Umständen für den Endanwender transparent erfolgen muss. Dies ermöglicht WASP-Anbietern, ihre Software regelmäßig und in kleinen Schritten zu aktualisieren und auf einen aufwendigen Roll-out einer stark angepassten Softwareversion – wie bei klassischer Software üblich – zu verzichten.

Als zusätzliche Services, die WASP-Anbieter neben der reinen Bereitstellung der Software erbringen können, sind typische Serviceleistungen, wie beispielsweise Schulungen oder Beratungsleistungen, zu nennen.

Die *Leistungsvereinbarung* (Phase 2) erfolgt in der Regel in Form einer Anmeldung, bei der die relevanten Informationen zur Person resp. zum Unternehmen festgehalten werden.

Die *Leistungserbringung* (Phase 3) erfolgt durch den Zugriff und die Nutzung der Software durch die Kunden.

Eine Besonderheit im WASP-Umfeld stellt der Kostenverlauf dar. Während bei der Konzeption und initialen Entwicklung der Software ein hoher Kostenblock anfällt, sind die Kosten bei der Einrichtung von Kunden auf der Software und der eigentlichen Nutzung der Software durch den Kunden zu vernachlässigen.

Leistungsangebotsmodell

Das Leistungsangebotsmodell wird im WASP-Umfeld als Folge der Standardisierung der Software nur geringfügig ausgeprägt. Neben der reinen Bereitstellung der Software können ggf. weitere Dienstleistungen angeboten werden.

Distributionsmodell

Bei der Betrachtung der Distributionsmodelle im WASP-Umfeld können zahlreiche Varianten, je nach Branchenfokus, Produktfokus etc. festgestellt werden.

Besondere Formen der Distribution im WASP-Umfeld sind z. B. das kostenlose Zurverfügungstellen einer eingeschränkten Version der Software oder – wie am Beispiel der ERP-Software *actindo Pro* erkennbar – das Einräumen einer längeren, kostenfreien Testphase [18].

Finanzierungsmodell

Initial wird ein Anbieter keine Erlöse durch die Software generieren können und beispielsweise Finanzinvestoren einbinden oder andere Leistungen seines Leistungsportfolios für die Finanzierung heranziehen.

Erlösmodell

In Abbildung 3 wird ein systematischer Überblick über die möglichen Formen der Erlösgenerierung in internetbasierten Geschäftsmodellen gegeben. Der Top-down-Ansatz zur Aufteilung in direkte und indirekte sowie transaktionsabhängige und transaktionsunabhängige Erlösformen ermöglicht eine Einordnung der im WASP-Umfeld auffindbaren Erlösarten, welche im Folgenden anhand einiger Beispiele kurz skizziert werden.

	Direkte Erlöse	Indirekte Erlöse
Transaktionsabhängig	- Anteil am Geschäftsvolumen - Transaktionsgebühr - Anteil einer Einsparung - Erfolgsprovision	- Fixe Provision - Erfolgsprovision
Transaktionsunabhängig	- Einrichtungsgebühr - Grundgebühr/Mitgliedsgebühr - Datenvolumengebühr - Zeiteinheitsgebühr - Gebühr pro Zugang	- Bannerwerbung - Tunneling - Bundeling - Sponsoring

Abbildung 3: Systematisierung der Erlösformen [Sch04]

Transaktionsabhängige, direkte Erlöse

Das Unternehmen *commercetools* stellt mit *ON Demand GRID* eine Software für den Onlineversandhandel bereit, für deren Nutzung eine Transaktionsgebühr erhoben wird. Konkret bedeutet dies, dass ein Handelsunternehmen für jede von ihren Kunden abgegebene Bestellung eine fixe Transaktionsgebühr entrichten muss [14].

Transaktionsabhängige, indirekte Erlöse

Eine Form transaktionsabhängiger, indirekter Erlöse wäre beispielsweise die Finanzierung einer WASP-Sofware durch einen Dritten, um seinerseits Geschäftsaktivitäten auf Basis der tatsächlichen Inanspruchnahme der Software durch Nutzer initiieren zu können.

Ein Onlineshop kann vom Softwareanbieter ohne ein Entgelt zur Verfügung gestellt werden, sofern sich der Nutzer verpflichtet, bestimmte weitere Dienstleister, beispielsweise einen bestimmten Paketlieferservice, einzubinden. Letzter profitiert von der tatsächlichen Nutzung des Shops in Form von zusätzlich generierten Aufträgen, sodass er als Dritter eine transaktionsabhängige Gebühr für jede Sendung an den Shopanbieter zahlen würde.

Transaktionsunabhängige, direkte Erlöse

Das Unternehmen *myfactory* bietet mit der Software *myfactory.BusinessWorld* ein komplettes, modulbasiertes ERP-System in drei Produkten an [16]. Die Nutzung dieser Produkte durch Interessenten für eine bestimmte Userzahl erfolgt gegen einen monatlichen Paketpreis (Userzahl-Produkt-Kombination). Eine Besonderheit stellt in diesem Zusammenhang dar, dass in diesem Paketpreis ein festes userbasiertes Datenvolumen einkalkuliert ist und ein höheres Datenvolumen zusätzlich entgolten werden muss.

Transaktionsunabhängige, indirekte Erlöse

Mit *Goolge Apps* vertreibt *Google* eine Kommunikations- und Workgroup-Software für die private und kommerzielle Nutzung in zwei sich hinsichtlich des Leistungsumfangs unterscheidenden Produkten [17].

Das Google-Modell sieht vor, dass die Nutzer der Standardvariante kein Entgelt für die Nutzung von Google Apps entrichten müssen. Die Finanzierung von Google Apps erfolgt analog zu den

anderen freien Diensten von Google über Werbung. Darüber hinaus können User beispielsweise zusätzlichen Speicherplatz beziehen.

3.4 Vor- und Nachteile von WASP

In einer Betrachtung des WASP-Geschäftsmodells können Parallelen, aber auch klare Abweichungen zum ASP-Geschäftsmodell festgestellt werden. Dies motiviert eine Analyse der Stärken und Schwächen von WASP aus Kunden- und aus Anbietersicht.

Aus Kundensicht stellen die geringe Kapitalbindung, die Möglichkeit zur nachhaltigen Kostenreduktion und die Komplexitätsreduzierung durch Fremdvergabe der gängigen Betriebs- und Entwicklungsaufgaben einen großen Vorteil dar. Zudem können mit dem orts- und zeitunabhängigen Zugriff auf gemeinsame Dokumente flexible Formen der Zusammenarbeit etabliert werden.

Die Nachteile werden aus Kundensicht im Bereich Security und Akzeptanz gesehen. Die Weitergabe erfolgskritischer Daten wird mit einem Sicherheitsverlust gleichgesetzt. Darüber hinaus wird eine Fremdvergabe von IT-Leistungen als Kompetenzverlust der IT verstanden.

Die Ergebnisse spiegeln sich auch in einer Marktstudie wider, derzufolge zahlreiche der befragten Unternehmen keinen Einsatz von WASP planen [2]. Als weiterer Grund wird vor allem das Vorhandensein einer äquivalenten Lösung im Unternehmen genannt.

Aus Anbietersicht bietet WASP zahlreiche Vorteile, insbesondere ein hohes Maß an Standardisierung und damit einhergehend an Komplexitätsreduzierung. Zudem ist erwähnenswert, dass die Einrichtung neuer Kunden i. d. R. nur geringe Kosten verursacht.

Hindernisse resultieren insbesondere aus dem hohen Marktdruck, der erschwerten Kommunikation des tatsächlichen Mehrwerts und der einfacheren Austauschbar. Ferner ist als kritisch anzusehen, dass – bedingt durch das Scheitern von ASP – Interessenten WASP bereits im Vorfeld mit einer gewissen Skepsis betrachten könnten.

Einen systematischen Überblick über die Vor- und Nachteile von WASP liefert die folgende Darstellung.

	Vorteile	Nachteile
Kundensicht	- Geringe Kapitalbindung - Geringe Nutzungsgebühren - Planbare IT-Kosten - Geringe Komplexität - Flexibilität - Verfügbarkeit	- Mangelndes Vertrauen - Security - Unklarer Mehrwert - Erschwerte Integration - Geringe Individualisierung - Abhängigkeit vom Anbieter
Transaktionsunabhängig	- Geringe Komplexität - Hohe Skaleneffekte - Geringe Grenzkosten - Hoher Standardisierungsgrad	- Schlechte Reputation - Hoher Marktdruck - Erlösformen unklar - Einfache Austauschbarkeit - ggf. kurze Vertragsbindung

Abbildung 4: Vor- und Nachteile des WASP-Modells

4 Zusammenfassung und Ausblick

Mit WASP wird ein interessanter Ansatz als Alternative zum klassischen Lizenzmodell diskutiert, welcher auf den gewonnenen Erfahrungen eines ähnlichen Ansatzes, dem Application Service Providing, aufsetzt.

In einer Gegenüberstellung wurde festgestellt, dass ASP und WASP eine ähnliche Zielsetzung, nämlich die Bereitstellung von Software über das Internet, verfolgen und hierzu ähnliche Konzepte umsetzen. Einige der identifizierten Schwachstellen im ASP-Modell, beispielsweise die Überladung der Kundenlösungen durch individuelle Anpassungen und Entwicklungen oder die hohen Nutzungsgebühren, werden berücksichtigt und beseitigt. Andere Schwachstellen, beispielsweise das Thema Sicherheit und Vertrauen, können auch durch WASP nicht grundlegend behoben werden.

Die Analyse der Geschäftsmodelle hat eine hohe Flexibilität in den Teilmodellen aufgezeigt und verdeutlicht, dass auf Anbieterseite große Chancen im WASP-Markt liegen. WASP erscheint insbesondere für kleine Unternehmen interessant zu sein, da die üblicherweise hohen Investitionen in IT verringert werden und eine Nutzungsgebühr nur nach dem tatsächlichen Nutzungsgrad zu entrichten ist. Ferner besteht ein großer Vorteil in der Möglichkeit flexibler Formen der Zusammenarbeit innerhalb eines Unternehmens und unternehmensübergreifend.

Hinsichtlich der Erfolgschancen von WASP kann vermutet werden, dass im Consumer-Segment der Vorteil einer preisgünstigen resp. freien Softwarenutzung als entscheidender Vorteil wahrgenommen wird. Im Business-Segment sind im Gegensatz dazu weitere Maßnahmen einzuleiten, um die bestehenden Kritikpunkte zu beantworten und das Vertrauen der Interessenten und potenziellen Kunden zu gewinnen.

Wird eine kritische Marktgröße erreicht, so werden weitere Softwareanbieter auf den Trend reagieren und unter Umständen die etablierten und erfolgreichen Anbieter im klassischen Lizenzmodell dazu bewegen, über WASP als Alternative zum bestehenden Modell nachzudenken. Bis zu diesem Zeitpunkt bleibt kritisch zu hinterfragen, ob sich WASP dauerhaft durchsetzen kann.

Literaturverzeichnis

[AZ01] R. Alt, H. D. Zimmermann: Introduction to Special Section – Business Models, in: Electronic Markets 11. Jg., 2001, H. 1, S. 3-9.

[Sch04] A. C. Schwickert: Geschäftsmodelle im Electronic Business – Bestandsaufnahme und Relativierung, in: Arbeitspapiere Wirtschaftsinformatik, Nr. 2. Siegen 2004.

[Stä01] P. Stähler: Geschäftsmodelle in der digitalen Ökonomie: Merkmale, Strategien und Auswirkungen, Köln 2001.

[vGu07] A. von Guten: Mit Software als Service zum Erfolg, in: InfoWeek.ch, 3. Jg., 2007, H. 2, S. 35-38.

[VH07] G. Vossen, S. Hagemann: Unleashing Web 2.0. From Cencepts to Creativity, Amsterdam u. a. 2007.

[Wir07] B. W. Wirtz: Electronic Business. 2. Aufl., Wiesbaden 2001.

Verzeichnis von Webadressen

[1] Software-Lego gehört die Zukunft,
http://www.computerwoche.de/itevolution/news/584445/
Zuletzt abgerufen am 20.08.2007.

[2] «Software as a Service» – Mietsoftware ist für kleine Unternehmen (noch) kein Thema,
http://www.pressebox.de/pressemeldungen/microsoft-presseservice/boxid-90533.html
Zuletzt abgerufen am 20.08.2007.

[3] Gartner erwartet ein erhebliches Wachstum für SaaS,
http://www.computerwoche.de/nachrichten/598112/
Zuletzt abgerufen am 20.08.2007.

[4] Software as a Service wird zum Mainstream,
http://www.cio.de/knowledgecenter/ondemand/828137/index.html
Zuletzt Abgerufen am 20.08.2007

[5] Die Wiedergeburt von ASP: Software as a Service, http://www.crmmanager.de/magazin/artikel_708_die_wiedergeburt_von_asp_software_as_a_service.html
Zuletzt abgerufen am 20.08.2007.

[6] Der Weg von ASP zu SaaS, http://www.it-business.de/themenkanaele/produkte/software/allgemein/articles/64278/ Zuletzt abgerufen am 20.08.2007.

[7] SaaS 101: Digging through definitions. 2006, http://blogs.zdnet.com/BTL/?p=4051
Zuletzt abgerufen am 20.08.2007.

[8] Software, http://de.wikipedia.org/wiki/Software
Zuletzt abgerufen am 20.08.2007.

[9] Dienstleistung, http://de.wikipedia.org/wiki/Dienstleistung.
Zuletzt abgerufen am 20.08.2007.

[10] Seeking the true meaning of SaaS, http://blogs.zdnet.com/SAAS/?p=251
Zuletzt abgerufen am 20.08.2007.

[11] 30 Boxes | Online Calendar, http://30boxes.com
Zuletzt abgerufen am 20.08.2007.

[12] Num Sum – web spreadsheet,' http://numsum.com/
Zuletzt abgerufen am 20.08.2007.

[13] Marktführer bei CRM-Software – salesforce.com,
http://www.salesforce.com/de
Zuletzt abgerufen am 20.08.2007.

[14] commercetools – On Demand eCommerce Software,
http://www.commercetools.de/
Zuletzt Abgerufen am 20.08.2007.

[15] R. Roewekamp: Software as a Service – Eine Software für alle,
http://www.cio.de/knowledgecenter/crm/827914/
Zuletzt abgerufen am 20.08.2007.

[16] myfactory.com On Demand Business, http://www.myfactory.com/
Zuletzt Abgerufen am 20.08.2007

[17] Google Apps, http://www.google.com/a/
Zuletzt Abgerufen am 20.08.2007

[18] Einzigartige ERP-Software für Linux-Systeme spart Kosten im Mittelstand,
http://www.pressebox.de/pressemeldungen/actindo-gmbh/boxid-101117.html
Zuletzt Abgerufen am 20.08.2007

Monetisierung großer Datenmengen

Jan Lammers

1 Paid-Content – Wer ist bereit, dafür zu zahlen?

Die aktuelle Situation im Traffic-Ranking[1] signalisiert deutlich einen steigenden Trend hin zu reinen Web 2.0-Anwendungen mit dahinter liegenden riesigen Datenmengen an Content. Insbesondere das vielseitige, größtenteils kosten- und werbefreie und stets um innovativere Facetten erweiterte Angebot in Form von User-Generated Content zieht weltweit millionenfach User an, erhöht allerdings zugleich den Aufwand und die Kosten für die gewaltigen Programmier-, Datenhaltungs- und Traffic-Kosten. Ein erfolgreiches Wirtschaften trotz dieser bekannten Schwierigkeiten ist äußerst schwierig – allerdings möglich[2] – und das Thema Content im Web 2.0 für Investoren aktuell attraktiv.[3]

Diese Situation wirft Fragen nach den verwendeten Geschäftsmodellen der Content-Provider auf, die im Folgenden als eine Kategorie von E-Business-Geschäftsmodellen näher untersucht werden. Dazu wird auf den von TIMMERS geprägten Begriff des E-Business-Geschäftsmodells eingegangen. Ferner werden die einzelnen Bestandteile eines Geschäftsmodells, wie die Architektur der Wertschöpfung, der Nutzwert und das Erlösmodell, im Kontext von Content-Providern vorgestellt.

2 Definition E-Business-Geschäftsmodell

Der Begriff des Geschäftsmodells wurde bereits 1998 durch TIMMERS im Kontext des E-Business geprägt:

- *„An Architecture for the product, service and information flows, including a desription of the various business actors and their roles; and*
- *a description of the potential benefits for the various actors; and*
- *a description of the sources of revenues.“*[4]

Folglich ist unter einem Geschäftsmodell *„die modellhafte Beschreibung eines Geschäfts“*[5] zu verstehen. Konkret bedeutet dies, dass die Frage „Welche Leistung wird durch wen erstellt?“

---

1 Vgl. Alexa (2007): youtube.com auf Rang 4, myspace.com auf Platz 6, facebook.com auf Platz 12 der Alexa Top Pages.

2 Vgl. XING (2007): Das deutsche Business-Networking-Portal XING erwirtschaftete 2006 einen EBITDA von 1,4 Mio. €.

3 Vgl. Internetworld (2007): Von insgesamt 26 Mrd. $ Venture Capital, das im Jahr 2006 weltweit investiert wurde, wurden 844,4 Mio. $ für Web 2.0-Unternehmen gezahlt.

4 Timmers (1998), S. 3.

5 Stähler (2001), S. 41 f.

und der zu erwartende Nutzen beschrieben werden. Im Fokus der vorliegenden Arbeit steht dabei allerdings vor allem das Erlösmodell mit der Frage „Wodurch wird Geld verdient?“.

Analysekriterium	Frage
Architektur der Wertschöpfung	Welche Leistung wird durch wen erstellt?
Nutzwert (Value Proposition)	Welchen Nutzen stiftet das Unternehmen seinen Zielkunden in welchem Marktsegment?
Erlösmodell (Revenue Model)	Wodurch wird Geld verdient?

Abbildung 1: Analysekriterien von E-Business-Geschäftsmodellen[6]

Diese drei aufgeführten und im Folgenden näher erläuterten Analysekriterien stellen keinen Anspruch an eine vollständige Beschreibung von E-Business-Geschäftsmodellen, da sie neben der Generierung von indirekten und direkten Erlösen oftmals auch parallel für andere Zwecke innerhalb der Wertschöpfung genutzt werden (Marketing, Wissensmanagement, Beziehungsmanagement etc.).[7]

3 Besonderheiten der Wertschöpfung im Bereich Paid-Content

Die Architektur der Wertschöpfung muss aufzeigen, welche Stufen der Wertschöpfung involviert sind und welche wirtschaftlichen Agenten in welchen Rollen aufseiten der Beschaffung („*distribution channels*“) und andererseits auf der Konsumentenseite („*target customer and market*“) des Contents notwendig sind.[8]

Die Erstellung der Marktleistung respektive des Contents und deren Absatz am Markt kann durch Wertschöpfungsprozesse grundsätzlich mit der Sequenz „Input – Process – Output“ beschrieben werden.[9] Im Gefüge von E-Business-Geschäftsmodellen und im Besonderen von Content-Providern getriebenen Geschäftsmodellen ist diese Sequenz allerdings deutlich komplexer, wie die folgende Aussage von O’REILLY belegt: *„We have subscriptions, and direct sales to consumers and mediated retail sales, and advertising, and combinations of all of the above. We have people who make their money providing infrastructure or assistance in these models – ad agencies, printers, rack jobbers, distributors, retailers. It's a rich and complex environment.“*[10]

In Anlehnung an WIRTZ kann auf der Inputseite („*distribution channels*“) zwischen einem Beschaffungsmodell und einem darauf aufbauenden Leistungserstellungsmodell unterschieden werden.[11] Das Beschaffungsmodell beschreibt dabei, wie die Aufnahme und Vergütung von Content von externen Content-Providern geregelt ist. Das Leistungserstellungsmodell beschreibt die Aggregation der einzelnen Content-Quellen zu konsumfähigen Content, der den Zielkunden/-usern einen solchen Nutzwert suggeriert, dass sie entweder bereit sind, Werbeein-

---

6 Eigene Darstellung in Anlehnung an Timmers (1998), Döbler (2007), Vossen/Hagemann (2007), S. 242.
7 Vgl. Döbler (2007), S. 28; Vossen, Hagemann (2007), S. 242.
8 Vgl. Vossen, Hagemann (2007), S. 242; Stähler (2001), S. 65.
9 Vgl. Schwickert (2004), S. 10.
10 Vgl. O’Reilly (2004).
11 Vgl. Wirtz (2006), S. 90 f.

blendungen während des Konsums zu ertragen, den Content direkt oder indirekt über substituierende Endgeräte oder Dienste zu bezahlen oder diesen durch Spenden zu sponsern.

Dabei ist allerdings im Zusammenhang mit auf Web 2.0-basierenden Content-Anwendungen zu beachten, dass neben den klassischen Porterschen Wertschöpfungsprozessen der Beschaffung und Leistungserstellung die Nutzer und Endkunden direkt und meist ohne geldwerten Vorteil mit in die Wertschöpfung einbezogen werden können. Dies gelingt durch die konsequente Integration der Partizipation und der Selbstorganisation der User als ein Charakteristikum von Web 2.0-Anwendungen.[12] O'REILLY führt dieses als „User-Generated Content"[13] bezeichnete Verhalten von Usern, bei dem diese durch die Erstellung von Inhalten, Verlinkungen und Kommentierung ein Netzwerk von Daten generieren, als einen Erfolgsfaktor mit auf: *„leverage customer-self service and algorithmic data management to reach out to the entire web, to the edges and not just the center, to the long tail and not just the head."*[14] So gibt es nach O'REILLY zwei Arten, den User dazu zu bewegen, ohne geldwerten Vorteil Content zu erzeugen: entweder das Verhalten des Users im Netz aufzuzeichnen und zu analysieren (z. B. das Kaufverhalten von Usern zu Kaufempfehlungen für andere Kunden aufbereiten) oder ihn zum aktiven Mitwirken aufzufordern (z. B. Buchrezensionen auf Amazon zu schreiben).[15] Insbesondere die Aggregation von Daten (algorithmic data management) als zweiter Erfolgsfaktor wird nach Prognose von O'REILLY zunehmend an Bedeutung gewinnen: *„Für Unternehmen wird die Qualität ihrer Daten künftig entscheidend sein. Wir betreten eine Welt, in der sich Vorteile am Markt aus der Herrschaft über Datenquellen ergeben.*"[16]

4 Nutzwertstrategien zur Bindung von Usern

Der Nutzwert einer contentbasierten Web 2.0-Anwendung besteht darin, dem einzelnen User durch ein Alleinstellungsmerkmal (unique selling proposition) einen einzigartigen Kundennutzen zu generieren. Offensichtlich wird ein Alleinstellungsmerkmal bei „Destination-Sites".[17] Dies sind Webseiten, mit denen die User aufgrund der Popularität der Seite bestimmte Funktionen und im Kontext von Content lastigen Internetangeboten Informationen verbinden. So stellt Google die Destination-Site im Internet für Suchanfragen dar und YouTube ist auf dem besten Wege, dies für Videos im Internet zu werden. Der Nachteil einer Destination-Site besteht für Anbieter allerdings darin, sich als solche zu etablieren. Daher ist der Ansatz zur Etablierung einer Destination-Site a priori kein Geschäftsmodell, sondern erst die Reichweite einer bestehenden Destination-Site kann als Grundlage eines Geschäftsmodells genutzt werden, um durch Traffic-Deals, Promotion oder Kooperationen monetäre Gewinne zu erzielen.[18]

Um den Nutzwert für den User zu erhöhen und möglicherweise parallel daraus direkt eigene finanzielle Erfolge zu erzielen, bietet sich Unternehmen und deren Partnern allerdings auch der

---

12 Vgl. O'Reilly (2005).
13 Vgl. Wikipedia (2007a).
14 Vgl. O'Reilly (2005).
15 Vgl. O'Reilly (2006).
16 Vgl. O'Reilly (2006).
17 Vgl. Krigel (1999).
18 Vgl. Gugel, Müller (2007), S. 39.

Ansatz einer Go-wide-Strategie. Diese Strategie sieht eine gezielte Syndikation und Verbreitung des Contents vor. So können die User durch die Nutzung ihres präferierten Dienstes frei wählen, wo und wie sie die Inhalte konsumieren möchten. Dieser auf der Masse der Kooperationen basierende Ansatz, der für die enorme Reichweite des Contents sorgt, bietet für die Anbieter des Contents den Vorteil, dass die technischen Details und die Vermarktung der Plattform durch den Partner übernommen werden. Der Partner wiederum profitiert von dem professionell bereitgestellten Content und der Vermarktung durch den Anbieter.

Ein „Virtual-Pipe-Modell“[19] ist eine weitere Möglichkeit, ein Alleinstellungsmerkmal zu erzielen, indem Content und User fest an einen Dienst gekoppelt werden. Diese Bindung von Usern kann grundsätzlich durch ein Digital Rights Management (DRM), soziale Netzwerke, Plattformen mit besonderer Infrastruktur oder hochwertigem Content erfolgen.

So erfolgt die Alleinstellung durch ein DRM in erster Linie durch einen geschützten Kanal,[20] der sicherstellt, dass Inhalte nur über einen eigenen Player oder autorisierte Webseiten abzuspielen sind. Apple zeigt mit seinem iTunes Music Store eindrucksvoll, wie die Monetisierung großer Datenmengen über das Internet funktioniert. So kontrolliert Apple nahezu die gesamte Wertschöpfungskette, angefangen bei ihrem proprietären Shopsystem, dem damit verbundene DRM, dem dafür benötigten Player (iTunes Music Player) und dem Endgerät (iPod) zum Abspielen des Contents. Durch diese starken Abhängigkeiten ist Apple in der Lage, den Content ohne Aufschlag von hohen Gewinnspannen anzubieten, da dieser durch den finanziellen Gewinn, der durch den Verkauf der Abspielgeräte erzielt wird, subventioniert wird. Dieses Beispiel zeigt auch, dass nicht zwangsläufig durch die Bereitstellung von Content gewinnbringende Erlöse erzielt werden müssen, sondern dass dies auch durch den Content substituierende Endgeräte oder Partnerschaften mit Herstellern von innovativen Endgeräten geschehen kann, auf denen die Nutzung des Contents dann für einen besonderen Kundennutzen sorgt.

Durch soziale Netze können User ebenfalls in eine Virtual-Pipe gebunden werden. Insbesondere durch die soziale Bindung des Users an die Plattform in Form von „Freunden“, persönlichen Bildern und Videos, Informationen, Beiträgen und Kommentaren erreichen die Anbieter einer solchen Plattform eine gewisse Abhängigkeit des Users, da dieser keine Möglichkeit hat, seinen Content und seine Vernetzungen auf eine andere Plattform zu transferieren. Dadurch wird einerseits der User für auf die Plattform geschaltete Werbung sensibilisiert, zum anderen kann der Anbieter Erlöse durch den kostenpflichtigen Zugang zu der Plattform oder durch den Verkauf von aus der Plattform heraus aggregierten Daten oder durch den Verkauf der Plattform erzielen.[21]

Durch die Bereitstellung einer Infrastruktur zum Ablegen, Bearbeiten und Präsentieren von datenintensivem Content auf einer Webseite kann ferner auch eine Destination-Site oder Virtual-Pipe etabliert werden. Flickr oder Photobucket sind prominente Beispiele, die zeigen, dass die Bereitstellung einer Möglichkeit zur kostenlosen Ablage und Präsentation von Content schnell hohe Netzwerkeffekte erzielt und dadurch den Nutzwert signifikant erhöht. Gewinnmöglichkei-

---

19 Vgl. Kessler (2006).

20 MUSSER und O'REILLY sprechen in diesem Zusammenhang von einer „Control Strategy“, vgl. Vossen, Hagemann (2007), S. 253.

21 Vgl. Kapitel 2.3 Erlösmodelle (Revenue Model) von E-Business-Geschäftsmodellen.

ten bei erfolgreicher Etablierung sind dann in der gleichen Form wie bei sozialen Netzen gegeben.

Eine weitere Form, einen einzigartigen Kundennutzen zu vermarkten, besteht durch das Binden von hochwertigen Inhalten an die Plattform mit teilweise exklusivem Zugang.[22] So ist/war der Zutritt zu Plattformen wie OpenBC/XING oder Googlemail nach vorheriger Einladung durch ein bereits registriertes Mitglied möglich. Die Registrierung auf der Plattform kostet oftmals kein Geld, jedoch erzielt der Anbieter durch das Verkaufen von Premium-Content an den Nutzer Erlöse. Gemein haben diese Möglichkeiten zur Etablierung einer Destination-Site oder Virtual-Pipe die Kombination der von MUSSER/O'REILLY aufgeführten Strategien:[23]

- Creation-Strategie: Generierung von Daten oder Content, der nur aufwendig reproduziert werden kann und der quasi in Monopolstellung angeboten werden kann. Als Beispiel kann hier Navteq aufgeführt werden, deren Kartenmaterial die Grundlage für mobile PDA-basierte Navigationslösungen darstellt.
- Access-Strategie: Zugang zu Content, der oftmals nach der Creation-Strategie geschaffen wurde, anhand unterschiedlicher Preismodelle und Käuferschichten.

Ein wesentlicher Erfolgsfaktor für die Nachhaltigkeit von generierten Nutzwerten sind Netzwerkeffekte. Nach WIRTZ sind zu erzielende Netzwerkeffekte der wichtigste Erfolgsfaktor für ein erfolgreiches Geschäftsmodell und parallel eine Markteintrittsbarriere für Konkurrenten.[24] Netzwerkeffekte bezeichnen das Phänomen, dass Nutzen und Nutzerzahlen sich gegenseitig positiv beeinflussen, was gegeben ist, wenn ein hoher Nutzwert der Content anbietenden Webseite viele Nutzer anzieht und dadurch den Nutzwert der Webseite weiter erhöht. Dieser mit der Awareness-Theorie („*an understanding of the activities of others, which provides a context for your own activity*")[25] zu erklärende, sich selber verstärkende Sog an Neuanmeldungen oder Content-Generierung ist Voraussetzung für die Etablierung einer Destination-Site oder Virtual-Pipe, die dann durch die im folgenden Kapitel aufgezeigten Erlösmodelle monetisiert werden kann.

5 Erlösmodelle (Revenue Model) von E-Business-Geschäftsmodellen

Das Erlösmodell gibt Aufschluss über den Wert und die Nachhaltigkeit des Geschäftsmodells und beantwortet die Frage „Wodurch wird Geld verdient?".[26] Einem Erlösmodel liegen dabei unterschiedliche Erlösmöglichkeiten zugrunde, die für sich alleine oder im Verbund genutzt werden können. Einen Überblick über die verschieden Formen gibt die nachfolgende Übersicht nach SCHWICKERT – systematisiert nach direkter und indirekter Erlösgenerierung sowie der Abhängigkeit des Erlöses von der Art der Transaktion.[27]

---

22 Vgl. Gugel, Müller (2007), S. 42 und Döbler (2007), S. 29.
23 Vgl. Vossen, Hagemann (2007), S. 252 f.
24 Vgl. Wirtz (2006), S. 584.
25 Vgl. Dourish, Bellotti (1992), S. 107-114.
26 Vgl. Vossen, Hagemann (2007), S. 242.
27 Vgl. Schwickert (2004), S. 11 f.; Wirtz (2001), S. 90 f.

	Direkte Erlösgenerierung Zahlungen von Nutzern der Marktleistung	**Indirekte Erlösgenerierung** Zahlungen von Dritten
Transaktionsabhängig	Anteil des Geschäftsvolumens Gebühr je Transaktion Anteil einer Einsparung Erfolgsprovision	Fixprovision Erfolgsprovision
Transaktionsunabhängig	Einrichtungsgebühr Grundgebühr Mitgliedsgebühr Datenvolumengebühr Gebühr je Zugang Gebühr je Zeiteinheit	Bannerwerbung Tunneling Bundeling Sponsoring

Quelle: Wirtz (2001), S. 91.

Abbildung 2: Systematisierung der Erlösformen von E-Business-Geschäftsmodellen

Aktuell dominieren noch transaktionsunabhängige indirekte Erlöse in Form von Werbung und Sponsoring die Monetisierung von Content.[28] So beschreibt Harry Motro, CEO von Infoseek, den gegenwärtigen Umsatz zu 2/3 aus Werbung, 1/3 aus Sponsoring und einem verschwindenden Rest aus Transaktionserlösen.[29] KRAUSE/BÜSCHING sprechen der Umsatzsäule Werbung/Sponsoring insgesamt ca. 80 % der Einnahmen zu.[30] Jedoch werden insbesondere durch die aktuelle Entwicklung hin zur individuellen Anpassung des Contents an den Nutzer die Möglichkeiten durch Paid-Content und Content Syndication sowie Data Mining an Potenzial gewinnen. Im Folgenden werden die einzelnen Möglichkeiten zur Erlösgenerierung kurz vorgestellt.

5.1 Erlöse durch Werbung

Anbieter von Webseiten oder Portalen haben grundsätzlich die Möglichkeit, interessierten Partnern kostenpflichtige Werbeflächen auf ihren Webseiten zu verkaufen. Die Darstellung erfolgt durch Werbebanner, Textlinks, Split Screens oder Interstitials. Die Bezahlung erfolgt dann transaktionsabhängig über Klickraten oder unabhängig über periodische Fixgebühren. Diese traditionellen Formen der Werbung im Internet setzen allerdings meist voraus, dass der Anbieter der Seite einen entsprechenden Nutzwert seiner Seite in Form einer Destination-Site oder Virtual-Pipe etabliert hat,[31] da eine Plattform erst ab einer bestimmten Größe für lukrative und zahlungskräftige Werbepartner interessant wird. In diesem Zusammenhang ist die Generierung von Content als ein Mittel zum Zweck hin zu einer stark frequentierten Webseite zu sehen.

Einen anderen Weg, anhand von Content durch Werbeeinnahmen Erlöse zu erzielen, demonstriert Google mit seinen Diensten AdWords und AdSense.[32] Bei dem Dienst AdWords bietet Google potenziellen Käufern den zeitlich befristeten Erwerb von einzelnen Suchbegriffen (AdWords) an, die in engem Zusammenhang mit den zu bewerbenden Produkten stehen. Für diese Produkte werden dann bei einer Suche nach entsprechenden Schlüsselwörtern Werbelinks für den User eingeblendet. Im Falle eines Klicks des Users auf einen solchen Link hat dann der

28 Vgl. Wirtz (2006), S. 589.
29 Vgl. Burg (2007).
30 Vgl. Krause, Büsching (2002), S. 1.
31 Vgl. Kapitel 4.
32 Vgl. für eine ausführliche Darstellung der Methoden Vossen, Hagemann (2007), S. 246 f.

Werbeträger eine Cost-per-Click (CPC)-Gebühr an Google zu zahlen. Um einerseits den Werbetreibenden ein faires System zu bieten, aber parallel die Erlöse aus den CPC-Gebühren zu optimieren, muss eine Statistik über die „click-through"-Raten geführt werden, d. h. die Werbeangebote, die von Usern erfolgreich durch einen Werbelink aufgerufen wurden. Dies ist damit zu begründen, dass Zahlungen für AdWords und insbesondere für CPC-Gebühren auf Micropayments[33] basieren, die erst durch sehr hohe Transaktionsvolumen signifikante Erlöse erzielen. Diese Statistik, die ein riesiges Datenvolumen umfasst, erzielt daher nicht auf direktem Weg Erlöse für Google, sondern optimiert das Transaktionsvolumen der CPC-Gebühren und generiert somit indirekt Erlöse, da neben dem Erweb der AdWords vor allem der Rang in der beschrieben Statistik das Kriterium zur Rangfolge der Werbeplatzierung ist. Da der AdSense-Dienst als eine Art Go-wide-Strategie des AdWord-Dienstes angesehen werden kann, ist dort die Wirkung Statistik und damit von Content noch um ein Vielfaches höher einzuschätzen, da hier zudem ein Filtern von verdächtigen Klicks, einerseits von Konkurrenten des Werbetreibenden und andererseits von Anbietern des AdSense-Dienstes zur Erhöhung ihrer Provisionsgebühr, stattfinden muss.

Insbesondere in sozialen Netzwerken ist zudem eine ureigene Problematik der Werbung umkehrbar. So müssen Werbetreibende in den klassischen Medien abhängig von der Dauer, der Sendezeit sowie der Reichweite des Werbung sendenden Mediums Gebühren zahlen, ohne sicher zu sein, den anvisierten Zielkunden zu erreichen. Das Problem kann dadurch gelöst werden, dass im Internet durch Auswerten von Userverhalten und Userprofilen Werbung zielgruppenspezifisch geschaltet und erfolgsabhängig über CPC-Gebühren verrechnet werden kann. Dies gelingt in sozialen Netzwerken so erfolgreich, da User aufgrund ihrer eigenen gewollten starken Positionierung im sozialen Netzwerk viel von ihrer Persönlichkeit in Profilen abspeichern und freigeben. Indes ist zu beachten, dass insbesondere der Web 2.0-Nutzer sehr sensibel auf Content-Seiten geschaltete Werbung reagiert und dies als Wechselgrund zu anderen ähnlichen werbefreien Anbietern dienen kann, wenn kein entsprechender Nutzwert der Seite etabliert worden ist.

5.2 Erlöse durch Paid-Content

Insbesondere Anbieter mit großen Datenmengen an Content können diese durch Paid-Content monetisieren. Dies geschieht durch Erheben einer Nutzungsgebühr für bestimmte Inhalte der Seite entweder transaktionsabhängig, beispielsweise beim Herunterladen (z. B. Song bei iTunes etc.) oder Benutzen (z. B. Ansehen eines Live-Streams) von Content, oder transaktionsunabhängig durch Zahlung einer monatlichen Mitgliedsgebühr (z. B. bei XING). Paid-Content wird für die meisten Anbieter von Content-Seiten in naher Zukunft ein wichtiger Bestandteil ihres Geschäftsmodells werden, um trotz der enormen Kosten der Infrastruktur zur Darstellung des Contents weiter wirtschaftlich bestehen zu können.[34] Dies ist damit zu begründen, dass je länger die Nutzungsdauer des Contents ist, desto schwerer ist die Finanzierung durch Werbeeinnahmen, da Werbung einen Aktualitätsanspruch bedingt. Anbieter von Content in Form von E-Books, Videos, oder Onlinegames müssen zu großen Teilen auf Werbefinanzierung verzich-

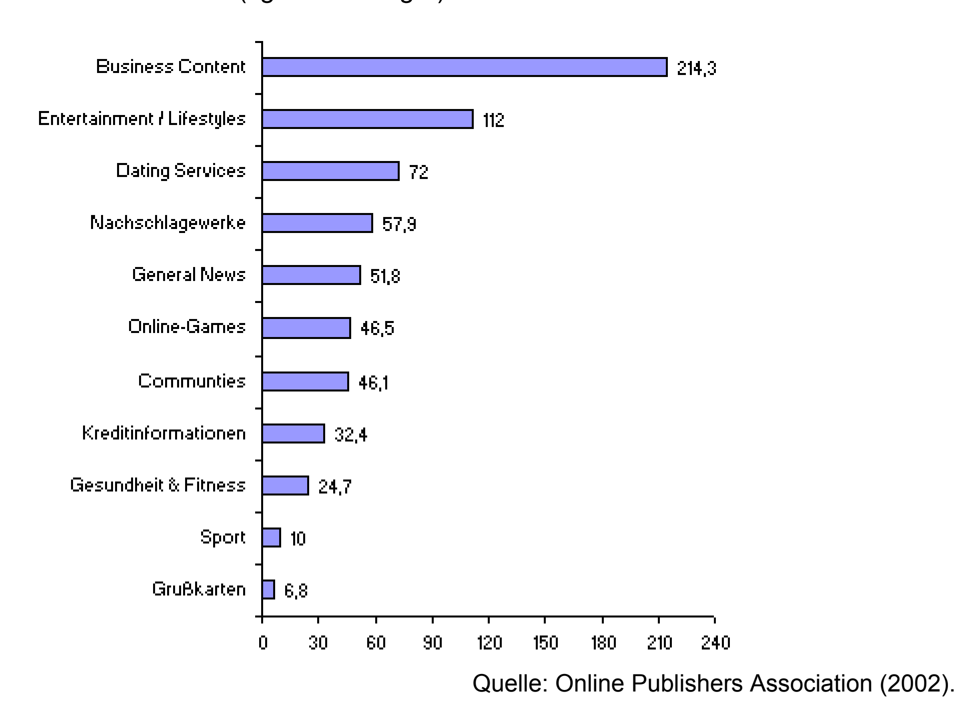
---

33 Nach herrschender Meinung fallen hierunter Beträge zwischen 0,01 und 5,00 €, die bei dem Erwerb eines digitalen Musikstückes, Zeitungsartikels oder Werbeangebots anfallen; vgl. Wikipedia (2007b).

34 Vgl. Meyer (2003).

ten, da der Nutzer andernfalls aufgrund seiner mangelnden Toleranz auf andere Angebote bis hin zu den klassischen Medien zurückgreifen wird.

Der Erfolg eines auf Paid-Content basierenden Geschäftsmodells ist davon abhängig, inwiefern durch Fokussierung auf Kundenbedürfnisse und Servicequalität für den einzelnen User ein solcher Nutzwert generiert wird, sodass er für den daraus resultierenden Mehrwert bereit ist Geld zu bezahlen.[35] Insbesondere in den Content lastigen Internetbereichen Onlinegaming (z. B. World of Warcraft, hattrick.org, Yahoo! „Games on Demand“), Musik (z. B. iTunes, Napster, Musicload) und Erotik als Ausprägung des Bereichs Entertainment sowie den Communitybereichen wie Businessportalen (XING) oder Partnerbörsen (z. B. neu.de) hat sich das Konzept des Paid-Contents bereits bewährt (vgl. Abbildung 3).



Quelle: Online Publishers Association (2002).

Abbildung 3: US-Online-Content-Ausgaben, geordnet nach Kategorien in Mio. $

KRAUSE/BÜSCHING nennen drei Geschäftsmodelle, die für den Verkauf von Premium-Content geeignet sind: das Vermittlermodell, das Vertriebsmodell, das Servicemodell.[36]

Bei dem Vermittlermodell verkauft der Content-Provider, der in der Folge als Verkäufer und Rechnungssteller agiert, kostenpflichtigen Inhalt oder Service an Kunden, die ihm durch ein Portal vermittelt wurden, das dafür eine Vermittlerprovision durch den Content-Provider erhält.

Bei dem Vertriebsmodell verkauft der Portalbetreiber Content direkt an den Kunden, den er zuvor gegen eine Lizenzgebühr oder einen Revenue Share bei dem Content-Provider erworben hat. Das Hosten der Inhalte kann bei dem Content-Provider oder dem Portalbetreiber erfolgen. Oftmals wird dabei der verkaufte Content noch mit dem Corporate Design des Portalbetreibers versehen (vgl. Apple’s iTunes Music Store), um so eine Virtual-Pipe zu etablieren.

---

35 Vgl. Krause, Büsching (2002), S. 2.

36 Vgl. im Folgenden Krause, Büsching (2002), S. 3.

Bei dem Servicemodell verkauft der Content-Provider direkt an den Kunden seine Inhalte, die Rechnungstellung erfolgt jedoch über einen Payment-Service-Provider, der gegen eine Gebühr als Clearingstelle agiert. Insbesondere wenn der Content-Provider nicht in der Lage ist, alternative gängige Zahlungsmöglichkeiten (Kreditkarte über SSL, Lastschriftverfahren, Virtual Money, e-Wallet etc.) anzubieten, ist ein solches Modell gebräuchlich.

Neben der Wahl des Geschäftsmodells für die Vermarktung von Paid-Contend ist das Pricing ein wesentlicher Bestandteil des Erlösmodells, da ein Erlösmodell ohne zeitliche und monetäre Quantifizierung ohne Aussage bleibt.[37] Dabei ist zu beachten, dass es aktuell noch sehr schwierig ist, Marktpreise für Content zu ermitteln, da bislang überwiegend Marktforschungsergebnisse über generelle Zahlungsbereitschaften für Content vorliegen,[38] allerdings für jede Leistung einer direkten oder indirekten Erlösgenerierung realistische Preismodelle inklusive Staffelungen/Rabatten/Bündelungen vorliegen müssen.[39] Dies kann durch Pricing-Strategien erfolgen, die durch fünf Kriterien zur Preisermittlung ergänzt werden (erzielter Nutzermehrwert, Preisniveau der Konkurrenz, Kosten plus Marge, Trial&Error) oder auf Basis von Marktforschungen:[40]

- „Penetration“: Strategie, den Markt durch einen anfangs niedrigen Einstiegspreis früh zu durchdringen und zu entwickeln und durch die Erzielung von Netzwerkeffekten eine Virtual-Pipe oder Destination-Site zu etablieren. Diese Strategie wird von einem Großteil der Anbieter von Content bevorzugt, da User aus der von kostenlosem Content geprägten Historie vornehmlich preissensibel reagieren und dementsprechend niedrige Preise hohe Marktanteile versprechen. Risiken birgt diese Strategie insofern, dass die Amortisierungszeit von den oftmals hohen Entwicklungs- und Traffic-Kosten sehr langwierig ist. Auch können spätere Preiserhöhungen meist nur sehr schwer durchgesetzt werden, da die User sich an das anfangs gestellte Preisniveau gewöhnt haben.
- „Skimming“: Strategie, die versucht durch einen hohen Anfangspreis früh eine Amortisierung der Entwicklungs- und Traffic-Kosten zu erreichen. Mit steigender Markterschließung und Wettbewerbsdruck wird dann der Preis stetig auf ein mittleres Niveau gesenkt, um dann zusätzlich breitere, aber preissensiblere Käuferschichten zu gewinnen. Erfolgreich ist eine solche Strategie dann, wenn der hohe Preis aufgrund einer hohen Content-Qualität oder Exklusivität[41] aus Sicht des Nutzers gerechtfertig ist. Das Risiko dieser Strategie liegt im Ausbleiben von Netzwerkeffekten aufgrund eines zu hoch gewählten Anfangspreises.
- „Teasing“ in Kombination mit „Trial&Error“: Strategie, die durch einen niedrigen Anfangspreis, der nicht als Kaufhindernis angesehen wird, anfangs eine hohe Probierrate erzielt. Diese Strategie eignet sich lediglich für neue auf Content basierende Webangebote und in erster Linie für prototypische Anwendungen, da durch Preisschwankungen nach oben hauptsächlich User abgeschreckt werden und dann keine Netzwerkeffekte zu erzielen sind.

---

[37] Vgl. Schwickert (2004), S. 12.
[38] Vgl. Reibnitz, Rautenberg, Schweiger (2003), S. 1.
[39] Vgl. Schwickert (2004), S. 12.
[40] Vgl. Reibnitz, Rautenberg, Schweiger (2003), S. 1.
[41] Vgl. Kapitel 4.

5.3 Erlöse durch Content-Syndication sowie Data Mining

Unter Content-Syndication ist die Verbindung von Inhalten verschiedener Webseiten zu verstehen.[42] Insbesondere Webseiten oder Portale, die ihr Angebot mit Business und Branchen relevanten Informationen aufwerten wollen, sind für Content-Provider unter dem Gesichtspunkt von Content-Syndication potenzielle Einnahmequellen. So können im Kontext von Content-Syndication auch eigene Inhalte gegen Lizenzgebühren an Dritte weitergegeben werden. Beispielsweise verbindet housingmaps.com den Service von GoogleMaps und Craiglist, einem großen Onlinekleinanzeigenmarkt für Immobilien, in einem Mash-Up, an dem GoogleMaps und Craiglist wahrscheinlich über Lizenzgebühren beteiligt werden.[43]

Der Content-Provider agiert im Kontext von Data Mining oftmals als „*Information Broker*“[44] und nutzt ein Geschäftsmodell, welches von RAPPA als „*Infomediary Model*“ bezeichnet wird.[45] Dabei sammelt der Content-Provider Daten über Kunden und ihr Verhalten im Web und unterstützt gegen Gebühren mit diesen dann aufbereiteten aggregierten Daten potenzielle Verkäufer bei dem Verständnis über einen Markt. Ein prominentes Beispiel ist DoubleClick, ein Netzwerk für Werbezwecke, das Unternehmen, Marketingfirmen und Content-Anbieter zusammenbringt, um erfolgreich digitale Werbung zu schalten.[46] Zu nennen sind an dieser Stelle allerdings auch soziale Netzwerke, in denen Unternehmen teils sehr große Summen für den Zugang zu Content in Form von detaillierten Userprofilen zahlen. Als Beispiel sei hier der Verkauf des StudiVZ genannt, das für geschätzte 50 bis 100 Mio. € an den Holtzbrinck-Verlag verkauft wurde,[47] der dafür Zugang zu 1,5 Mio. Mitgliederprofilen erhielt.

5.4 Erlöse durch Spenden (Donations)

Dass Erlöse aufgrund freiwilliger Spenden von Privatpersonen und Unternehmen zu erzielen sind, demonstriert das Beispiel Wikipedia. So lassen sich anscheinend trotz monatlicher Betriebskosten in Höhe von 75.000 €[48] durchaus Gewinne erzielen. Ein solches Erlösmodell setzt allerdings unbedingt die erfolgreiche Etablierung einer Destination-Site (im Falle von Wikipedia oder vergleichbar Sourceforge) oder einer sehr großen Virtual-Pipe voraus und ist zudem stark auf kostenlos nutzbaren User-Generated Content angewiesen.

Literaturverzeichnis

Alexa: http://www.alexa.com/, 2007, Abrufdatum: 2007-08-20.
Burg, T. N.: Geschäftsmodelle für die Internet-Ökonomie, 2007, http://randgaenge.net/wp-content/uploads/burg_destinationsite.pdf, Abrufdatum: 2007-08-29.

---

42 Vgl. Wikipedia (2007c).
43 Vgl. O'Reilly (2005).
44 Vgl. Vossen, Hagemann (2007), S. 248.
45 Vgl. Rappa (2007).
46 Vgl. Rappa (2007a).
47 Vgl. Spiegel (2007).
48 Vgl. Wikipedia (2007d).

Döbler, T.: FAZIT-Studie – Marktchancen durch Social Software, Stuttgart 2007, http://www.fazit-forschung.de/fileadmin/_fazit-forschung/downloads/Marktstudie_SocSoft.pdf, Abrufdatum: 2007-08-18.

Dourish, P., Bellotti, V.: Awareness and Coordination in Shared Workspaces. Proceedings of CSCW 92.

Gugel, B., Müller H.: TV2.0 Whitepaper, 2007, http://www.gugelproductions.de/blog/wp-content/uploads/2007/07/tv20_gugel_mueller.pdf, Abrufdatum: 2007-08-17.

Internetworld: http://www.internetworld.de/news-single.html?&tx_ttnews %5Btt_news %5D=1073&cHash=1cc55f82b9, Abrufdatum: 2007-08-30.

Kessler, A.: Media 2.Uh-Oh Part 3: Virtual Pipes, 2006, http://www.andykessler.com/andy_kessler/2006/10/media_2uhoh_par_2.html, Abrufdatum: 2007-08-17.

Krause S., Büsching A.: Geschäftsmodelle für Portale mit kostenpflichtigen Diensten, 2002, http://www.contentmanager.de/magazin/artikel_166-print_geschaeftsmodelle_portale.html, Abrufdatum: 2007-09-01.

Krigel, B. L.: New buzz word: „Destination“ site, 1999 http://news.com.com/2100-1023-219883.html, Abrufdatum: 2007-09-01.

Meyer, A.: Paid-Content – Der Markt für Online Inhalte, 2003, http://www.contentmanager.de/magazin/artikel_284-print_paid_content.html, Abrufdatum: 2007-09-01.

O’Reilly, T.: Tim O'Reilly Interview, Part 2: Business Models & RSS2004, 2004, http://www.readwriteweb.com/archives/tim_oreilly_int_1.php, Abrufdatum: 2007-08-18.

O’Reilly, T.: What is he Web 2.0, 2005, http://www.oreilly.de/artikel/web20.html Abrufdatum: 2007-08-18.

O’Reilly, T.: Neue Wetten in Silicon Valley, Die Zeit, 2006, http://images.zeit.de/text/2006/39/Interv_Reilly, Abrufdatum: 2007-08-18.

Vossen, G.; Hagemann, S.: Unleashing Web 2.0. From Cencepts to Creativity. Amsterdam u. a., 2007.

Rappa, M.: Business Models on the Web, 2007 http://digitalenterprise.org/models/models.html, Abrufdatum: 2007-08-18.

Rappa, M.: Case Studie: DoubleClick, 2007a http://digitalenterprise.org/cases/doubleclick.html, Abrufdatum: 2007-08-18.

v. Reibnitz, A.; Rautenberg, A.; Schwaiger, M.: Pricing von Paid-Content – Vorgehen bei der Preisfindung und gegenwärtiges Preisniveau, 2003, http://www.contentmanager.de/magazin/artikel_380-print_pricing_paid_content.html, Abrufdatum: 2007-09-01.

Schwickert, A. C.: Geschäftsmodelle im Electronic Business – Bestandsaufnahme und Relativierung, Arbeitspapiere Wirtschaftsinformatik 2004, http://www.econbiz.de/archiv/gi/ugi/winformatik/geschaeftsmodelle_eb.pdf, Abrufdatum: 2007-08-18.

Spiegel: Holtzbrinck schnappt sich StudiVZ, 2007, http://www.spiegel.de/netzwelt/web/0,1518,457536,00.html, Abrufdatum: 2007-08-25.

Stähler, P.: Geschäftsmodelle in der digitalen Ökonomie: Merkmale, Strategien und Auswirkungen. Lohmar, Köln 2001.

Timmers, P.: Business Models for Electronic Markets in:Gadient, Yves; Schmid, Beat F.; Selz, Dorian: EM – Electronic Commerce in Europe. EM – Electronic Markets, Vol. 8, No. 2, 07/98, http://www.electronicmarkets.org/modules/pub/download.php?id=electronicmarkets-183&user=&pass=, Abrufdatum: 2007-08-18.

Wikipedia: User-Generated Content 2007a http://de.wikipedia.org/wiki/User_Generated_Content, Abrufdatum: 2007-08-25.

Wikipedia: Micropayment 2007b
http://de.wikipedia.org/wiki/Micropayment, Abrufdatum: 2007-08-31.
Wikipedia: Content-Syndication 2007c
http://de.wikipedia.org/wiki/Content-Syndication, Abrufdatum: 2007-08-25.
Wikipedia: Wikipedia Finanzierung 2007c
http://de.wikipedia.org/wiki/Wikipedia#Finanzierung, Abrufdatum: 2007-08-29.
Wirtz, B. W.: Electronic Business. 2. Aufl., Wiesbaden 2001.
XING: http://corporate.xing.com/fileadmin/image_archive/ir_OPEN_Business_Club_
AG_Geschaeftsbericht_2006_deutsch.pdf, Abrufdatum: 2007-08-20.

Second Life

Gereon Strauch

1 Einleitung

Selten hat ein Phänomen der Internetökonomie die Ansichten so polarisiert, wie die virtuelle Welt Second Life in den letzten Monaten. Neben den unglaublichen Zuwachsraten und Geldmengen, die im virtuellen Raum bewegt werden, ist es vor allem das Konzept eines Metaversums, das stellvertretend ein zweites Leben in der Virtualität erlaubt, das schon immer eine starke Faszination auf die Menschen ausübt. Neben verschiedenen Aufbereitungen in der Literatur, wie z. B. *William Gibson* in seinem Buch *Neuromancer*[1] oder der literarischen Vorlage von Second Life, dem Roman *Snow Crash* von *Neal Stephenson*, wurde auch immer wieder der Gedanke der virtuellen Gemeinschaft praktisch umzusetzen versucht.[2] Second Life bedient diese Faszination – über den geschickt gewählten Namen hinaus – besonders durch das Angebot einer dreidimensionalen Welt, die die Anwender selber gestalten können. Neben einem anfänglichen Hype mehren sich in der letzten Zeit die kritischen Stimmen, die über Stabilitäts- und Performanceprobleme, Kriminalität und Gefährdung der Nutzer, Rechtsprobleme und ein überfordertes Management berichten. Stagnierende und rückläufige Nutzerzahlen, insbesondere der Rückzug zahlender Kunden,[3] alarmieren die Analysten, die zuvor in Second Life noch das Internet der Zukunft vermuteten. Derartige Hype-Zyklen sind zwar anscheinend unabdingbare Begleiterscheinungen der Berichterstattung über internetbasierte Phänomene, aber selten verlief eine Entwicklung so rasant wie hier. Das Konzept einer Nutzbarmachung der Erträge von User-Generated Content für eben die Anwender durch eine Web 2.0-basierte Wirtschaft aus der Sicht der Wirtschaftswissenschaft ist dringend überfällig. Dieser Beitrag hat den Fokus, wirtschaftliche Implikationen und Möglichkeiten, die sich durch Second Life ergeben, zu beleuchten. Dazu werden zunächst die technischen Aspekte erläutert. Das Web 2.0 wird vor allem auch als soziales Phänomen beschrieben. Anhand wichtiger Fragestellungen und Probleme, die einer unternehmerischen Tätigkeit auf Second Life entgegenstehen können, werden dann wichtige gesellschaftliche Institutionen und ihre bisherige Umsetzung oder noch notwendige Implementierung in Second Life angerissen, um den Hintergrund einer fundierten wirtschaftswissenschaftlichen Betrachtungen darstellen zu können. Im Folgenden werden dann die wirtschaftlichen Möglichkeiten auf und mit Second Life analysiert. Dazu wird ein Ordnungsrahmen für Geschäftsmodelle im Web 2.0 eingeführt. Dabei sollen zunächst die real stattfindenden und potenziellen wertschöpfenden Tätigkeiten im Rahmen von Geschäftsmodellen mit Second Life beschrieben und deren prinzipielle Ertragsaussichten betrachtet werden. Die untersuchten Geschäftsmodelle sind in ihrer Entwicklung von der des Second Lifes abhängig. Daher wird anschließend Second Life vor dem Hintergrund der Prinzipien der Web 2.0-Ökonomie selber als Geschäftsmodell beleuchtet. Damit ist nicht nur eine Beurteilung des Geschäftsmodells Second Life, sondern auch der Zukunftsfähigkeit von Geschäftsmodellen auf Second Life möglich. Ab-

---

[1] Vgl. Niemeyer (1998).

[2] Schon 1985 vermutlich zum ersten Mal mit „The Well“, http://www.thewell.com.

[3] http://pressetext.com/pte.mc?pte=070904003 [11-09-09].

schließend werden aus den gewonnenen Erkenntnissen Gestaltungsempfehlungen für Second Life formuliert, die sich auch auf ähnlich gestaltete virtuelle Welten übertragen lassen.

2 Technik

Second Life ist die dreidimensionale Simulation einer virtuellen Welt, an der jeder der Nutzer, die in Second Life als Residents (Bewohner) bezeichnet werden, über einen Avatar teilnehmen kann. Dabei können die Residents sich innerhalb dieser virtuellen Welt bewegen, mit anderen Usern kommunizieren und sonstig aktiv werden kann. Der Anbieter dieses Systems ist das Unternehmen Linden Lab.[4] Die Besonderheit an diesem System ist, dass die Anwender selber Land erwerben und virtuelle Güter erstellen können. Auch die Avatare sind weitgehend vollständig selbst zu gestalten. Um einen Avatar erstellen und damit an dieser Community partizipieren zu können, ist lediglich die Freischaltung eines Accounts beim Betreiber Linden Lab und die Installation eines relativ kleinen Clients auf dem eigenen Rechner notwendig. Der Anwender – oder in der Diktion von Second Life Einwohner oder Resident – wird bei seiner Anmeldung aufgefordert, sich einen Avatar aus einer Liste von verschiedenen menschlichen oder tierähnlichen Avataren mit Menschengestalt, so genannten Furries, auszuwählen. Der Avatar ist frei anpassbar, verschiedene Aussehen und Kleidungen kann der Anwender in einem Inventar speichern. Hier werden auch verschiedene Bewegungsmuster für den Avatar (Poses) hinterlegt.

Alle Objekte in Second Life bestehen aus einfachen Grundobjekten, so genannten Primes (von primitiv), die modifiziert, verbunden und mit Texturen belegt werden können. Diese Objekte können mit der Linden Skript Language (LSL), einer an C++ angelehnten Skriptsprache, mit Funktionalität versehen werden. So können die Residents ihre Welt selber gestalten. Damit wurden über 99 % der bestehenden Infrastruktur auf Second Life von den Anwendern selber erzeugt. Die Objekte werden über ein Rechtekonzept verwaltet, um Missbrauch, Vandalismus und Urheberrechtsverletzungen zu vermeiden. Im Second Life Grid ist die Havok Physik Engine implementiert worden, damit Objekte auch mit einem realistischen Verhalten ausgestattet werden können.[5] Da Second Life XML-RPC und HTTP-Requests unterstützt, ist eine Anbindung nach externen Anwendungen möglich.[6]

Das Land ist in dem so genannten Grid, der simulierten Welt von Second Life, in verschiedene Regionen aufgeteilt. Neben dem Mainland, das dem Betreiber Linden Labs gehört, gibt es Inseln in Privatbesitz, so genannte Sims. Diese umfassen eine Größe von je 65.536 m^2, die auf einem dedizierten physikalischen Server betrieben werden. Auf einer Insel beträgt die Höchstsumme an Prims 15.000. Entgegen dem Architekturkonzept anderer virtueller Onlinewelten werden die 3D-Daten nicht auf dem Rechner des Users vorgehalten, sondern über die Simulationsrechner des Grids. Damit ist es nicht nur möglich, die Clientsoftware klein zu halten, sondern auch den Usern zu erlauben, laufend Änderungen in das System einzubringen, die von allen anderen Usern in Echtzeit übernommen werden können. Daneben werden noch unterschiedliche Server zur Authentifizierung, für die Nutzerverwaltung, die räumliche Zuordnung etc. betrieben. Einen grundsätzlichen Überblick über die Systemarchitektur bietet Abbildung 1.

---

4 http://www.lindenlab.com/ [05-09-07].

5 http://www.havok.com/content/view/17/30/ [09-09-09].

6 Vgl. Moore (2007), S. 39 f.

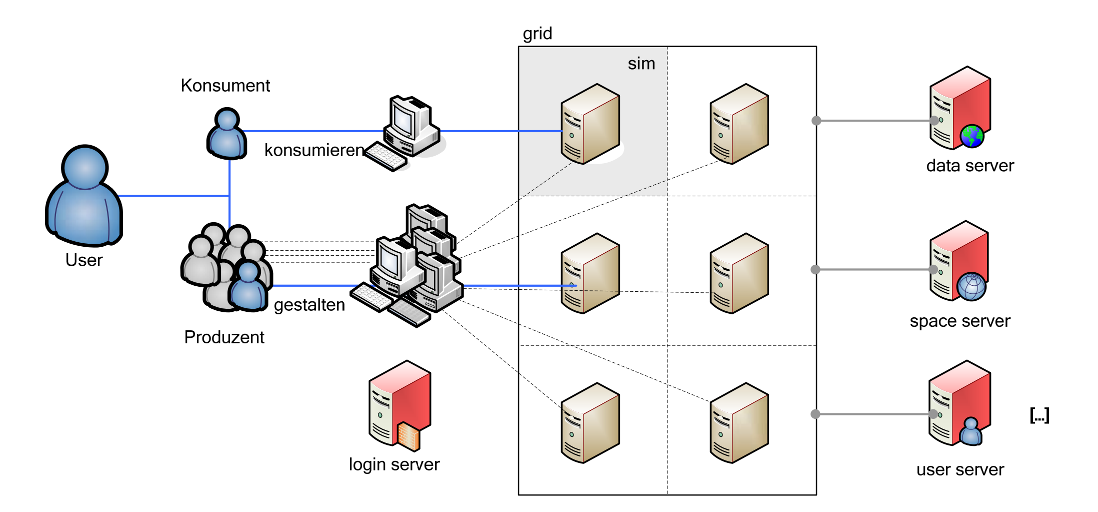


Abbildung 1: Systemarchitektur[7]

Weder das System noch der Client zeichnen sich durch eine besonders gute Usability oder Stabilität aus. Die Aktionsmöglichkeiten von Anwendern sind fast unbegrenzt. Zunächst ist die Gestaltung des Avatars über Kleidung, Skins, Form und Accessoires ein häufiges Betätigungsfeld. (Hier scheint der Avatareffekt zum Tragen zu kommen.) Daneben lässt sich das Verhalten des Avatars über verschiedene Animationen, so genannte Pose Sets, anpassen. Darüber hinaus können die Avatare an Angeboten anderer teilnehmen oder eigene erstellen. Dafür ist in der Regel Geld der Second Life eigenen Währung (Linden Dollar, LD) notwendig, das entweder von anderen Residents über eine Börse (LindeX) oder direkt bei Linden Labs zum ungefähren Durchschnittskurs bezogen werden kann.[8] Über die Börse ist es auch möglich, Linden Dollars zurückzutauschen.

3 Gesellschaft

3.1 Gesellschaftliche Dimension von Second Life

Die Kernidee von Web 2.0-Anwendungen, die Generierung der Inhalte durch Anwender, wird im Konstruktionsprinzip von Second Live auf die Spitze getrieben. Nicht nur Inhalte sind in Second Live durch die Anwender erstellbar, sondern auch die Art der Interaktion, die Anwendungen und Strukturen. Also lassen sich in Second Live Subcommunities für (fast) jeden Zweck erstellen. Dabei sind die unterschiedlichen Anwendungen zwar räumlich virtuell getrennt, was jedoch durch die Möglichkeiten zur Fortbewegung von Avataren keine Rolle spielt. In der Tat hat es noch keine derart anschauliche Umsetzung des Prinzips einer „flachen Welt“[9] gegeben. Dadurch, dass die Möglichkeiten der Interaktion und Gestaltung durch die Avatare annähernd unbegrenzt sind und Transaktionen durch die Einführung einer Währung begünstigt werden, ist sogar davon auszugehen, dass durch Second Life eine Art von Gesellschaft entstanden ist oder zumindest derzeit evolviert.

---

7 http://wiki.secondlife.com/wiki/Server_architecture [10-09-07].

8 http://www.second-life.com/de/linden_dollar.html [10-09-07].

9 Friedmann (2006).

Beispiele für die Entwicklung einer Art von Gesellschaft finden sich in den Aktivitäten, die dort stattfinden. Einerseits sind – wie im Internet üblich – verschiedene Interessengruppen zu finden, die ihre Subcommunities auf Second Life betreiben. Dabei haben sich – wie in einer Realgesellschaft – unterschiedliche gesellschaftliche Institutionen, z. B. zur Politik, Kultur und Wirtschaft, oft mit einem expliziten Bezug zum System Second Life gebildet. Daneben finden sich auch derartige Einrichtungen mit Bezug zu realen Entitäten, wie in der Politik. So hat z. B. Nicolas Sarkozy in Second Life Wahlkampf betrieben.[10] Im kulturellen Bereich finden sich auch vielfältige Aktivitäten, angefangen vom Fan, der seine Lieblingsband U2 auf einem virtuellen Wohltätigkeitskonzert auftreten lässt, bis hin zur bekannten realen Popgruppe Duran Duran, die – wie andere auch schon – ein Konzert in Second Life veranstaltet hat.

Diese „virtuelle" Gesellschaft hat auch schon eine Volkswirtschaft mit allen wesentlichen Elementen entwickelt, wie einem Arbeitsmarkt,[11] Unternehmen und auch Leistungsbeziehungen zu Volkswirtschaften außerhalb. Die Volkswirtschaft von Second Life wird im Jahr 2007 wahrscheinlich ein Bruttoinlandsprodukt (BIP) von über 700 Mio. $ erreichen und befindet sich damit schon in einer Größenordnung zwischen den realen Volkswirtschaften von Guyana und Dschibuti, also Ländern mit 500.000 bis 750.000 Einwohnern.[12] Damit gehen aber auch gesellschaftliche, rechtliche und wirtschaftliche Probleme einher. So ist z. B. mit einem derartigen Währungsregime, wie dieses in Second Life vorliegt, mit Inflation zu rechnen.[13] Außerdem ist, da durch die – zumindest partiell – konvertible Währung letztendlich reale Werte gehandelt werden, mit Auseinandersetzungen über Eigentumsverhältnisse zwischen den Akteuren zu rechnen.

3.2 Institutionelle Probleme

Mit der emergenten Entwicklung einer Gesellschaft gehen auch die entsprechenden Arten von Unkultur, Straftaten und anderen gesellschaftlichen Problemen solcher sozioökonomischer Systeme einher.[14] Dabei ist – wie sonst auch im Internet – vor allem auch damit zu rechnen, dass anderweitig sanktioniertes Verhalten in den scheinbar rechtsfreien und vor allem aber unübersichtlichen Raum virtueller Gemeinschaften ausweicht. So sind bei Second Life leider schon Fälle von virtueller Kinderpornografie beobachtet worden. Problematisch ist zudem, dass die Plattform auch zur Verabredung anderer Straftaten im realen Leben genutzt werden kann, weil die Kommunikationswege dort kaum überwacht werden können und das Grid jetzt schon derart groß und heterogen ist, dass eine Übersicht kaum noch erlangt werden kann. Durch das finanzielle Umtauschsystem ließen sich auf diesem Weg auch Gelder waschen. Daneben sind durchaus auch straf- und zivilrechtliche Aspekte von Handlungsweisen innerhalb von Second Life von Relevanz. Die Frage ist z. B. auch, wie sich Eigentums- und Persönlichkeitsrechte in Second Life schützen lassen. Dies betrifft einerseits die von realen Personen, die in Second

---

10 Vgl. Casati et al. (2007).

11 D. h., dass es auch für Residents die Möglichkeit gibt, mit unselbstständiger Arbeit auf Second Life Geld zu verdienen. Vgl. dazu Kap. 4.2.1.

12 http://de.wikipedia.org/wiki/Liste_der_L %C3 %A4nder_nach_Bruttoinlandsprodukt [09-09-09].

13 Vgl. Zimprich (2006). Dabei ist insbesondere das Wachstum interessant, da hier in einer realen Wirtschaft von einer Überhitzung auszugehen wäre: 2006 lag das BIP bei zirka 64 Mio. $, vgl. http://www.second-ife.com/de/linden_dollar.html [09-09-09]. Das Phänomen der Inflation wird schon seit Längerem auch in MMORPGs beobachtet und in diesem Kontext auch als Mudflation bezeichnet (nach der Bezeichnung der ersten Onlinerollenspiele, so genannten MUDs, und Inflation), vgl. http://de.wikipedia.org/wiki/Mudflation [09-09-09].

14 Vgl. Casati et al. (2007).

Life dargestellt werden (z. B. Prominente), aber auch die von Avataren. So können nach deutschem Recht auch Avatare eine Persönlichkeit haben.[15] Andere im virtuellen Leben vorstellbare Straftaten wären Diebstahl/Raub, Betrug und Computerbetrug, Nötigung/Erpressung, Hehlerei, Verletzung des persönlichen Lebens- und Geheimnisbereichs und Straftaten gegen die sexuelle Selbstbestimmung.

Erschwerend kommt dabei hinzu, dass – wie üblich im Internet – durch den Fakt der Deterritorialisierung die Frage nach dem zu verwendenden Recht nicht eindeutig geklärt ist. In Second Life sind dabei weder das Vorgehen noch die Konsequenzen in derartigen Fällen geklärt. Es bestehen so gut wie keine Institutionen, die derartige zwangsläufig auftretende Konflikte lösen helfen können. Second Life hat jedoch keine Institutionen in der Art von Regelungen oder Organisationen definiert, die derartige Konflikte lösen. Nur diese könnten die fundamentale Unsicherheit beseitigen, die solche Aktivitäten verursachen und so die Transaktionskosten soweit senken, dass auch weiterhin Transaktionen stattfinden.

Es ist bisher zu beobachten, dass sich Anbieter von Web 2.0 mit der Gestaltung dieser Institutionen recht schwer tun.[16] Einerseits ist eine vollständige Überwachung und Reglementierung des Userverhaltens zu aufwendig, andererseits reagieren die Nutzer bei den zumeist dann unangemessenen deterministischen Eingriffen sehr sensibel.[17] Es müssen daher institutionelle Arrangements geschaffen werden, bei denen die Selbstorganisation der Netzwerke zu einer Einhaltung externer Erfordernisse führt. Einerseits handeln sich mit einem derartigen Verhalten die Betreiber von Web 2.0-Plattformen selber rechtliche Probleme ein, da sie teilweise von den Gerichten für das (Fehl-)Verhalten von Nutzern in die Verantwortung genommen werden.[18] Ohne Institutionen, die diese Problemfelder regeln helfen, bedeutet dies aber fehlende Investitionssicherheit für Anwender auf der proprietären Plattform Second Life. Daher sind Institutionen in diesem Sinn vor allem wichtig für Second Life, da es in seiner weiteren Entwicklung auch auf die Investitionen von kommerziellen Anwendern angewiesen ist. Bei einem Scheitern dieser Plattform wären diese Investitionen allerdings verloren, damit wären Beiträge auch von kommerziellen Anbietern nicht mehr zu erwarten. Damit könnte sich also ein negativer Netzeffekt gegen langfristige Investitionen kommerzieller Akteure in Second Life ergeben. Für Linden Labs besteht an diesem Punkt also wesentlicher Handlungsbedarf.

Gerade in den emergenten, zur Evolution fähigen soziotechnischen Systemen unserer Gesellschaft haben sich differenzierte Institutionen herausgebildet (z. B. zur Unsicherheitsreduktion, d. h. Senkung der Transaktionskosten). Besonders problematisch ist in diesem Kontext, dass Second Life schon eine an die reale Welt gekoppelte Volkswirtschaft darstellt. An diesem Beispiel lässt sich gut die Problematik verdeutlichen: So nimmt Linden Labs z. B. keinerlei Notenbankfunktionalität bei der Sicherung der Konvertibilität wahr, sondern übernimmt bisher nur den Wechsel von $ in LD. Ein Rücktausch ist nur über die systemeigene Börse LindeX möglich, bei der im Endeffekt Residents miteinander handeln und – wie an einer Börse üblich – keinerlei

---

15 Vgl. Hopf, Braml (2007), S. 354.

16 Ein Beispiel ist das fehlende Vorgehen der Betreiber gegen extremistische Gruppierungen auf StudiVZ, vgl. Pham (2007).

17 Vgl. Lischkar (2007a), Lischkar (2007b).

18 Vgl. Lischkar (2007c). Eine sichere Überprüfung, wie vom deutschen Gesetzgeber (z. B. mit Einsatz einer geschlossenen Benutzergruppe i. S. v. § 4 Abs. 2 Satz 2 JMStV oder der Einsatz eines Jugendschutzprogramms i. S. v. §11 JMStV) gefordert, existiert damit aber bisher nicht; vgl. Krasemann (2006), S. 354.

Rücktauschgewähr zum gewünschten Kurs besteht. Durch fehlende Regelungen sind in bestimmten Bereichen der Wirtschaft von Second Life durchaus Probleme zu erwarten.[19]

Eine Regulierung ist damit unumgänglich. Die ausschließliche Institutionensetzung durch den Betreiber Linden Lab könnte von den Residents als diktatorisch empfunden werden und damit das eigene Geschäftsmodell gefährden, das Web 2.0-spezifisch auf der Beteiligung der Akteure beruht. Eine Beteiligung der User als dem „Volk von Second Life" ist wegen des mangelnden Reifegrades der Community und dem Fehlen zentraler Strukturen (z. B. zur Massenkommunikation) noch nicht realistisch. Dabei wären Vorgehensmodelle und Schiedsgerichte vorstellbar, ggf. Eskalationsregeln für die Überweisung an reale Instanzen. Dabei müsste auch aus technischer Sicht an der Schaffung und Institutionalisierung von Beweismöglichkeiten und -pflichten gearbeitet werden. Intern zu regelnde Institutionen – die teilweise im Einvernehmen mit den Usern zu regeln sind – wären z. B. eine Börsenaufsicht, Bankenrecht und -aufsicht, eine Notenbank (die Konvertibilität sicherstellt) und Regelungen zum Arbeitsrecht und Aufsichtsinstanzen.[20]

4 Ökonomie

4.1 Geschäftsmodelle und Second Life

4.1.1 Geschäftsmodelle im Web 2.0

Um rationale Entscheidungen über die Auswahl und Ausgestaltung von unternehmerischen Aktivitäten in evolvierenden Umfeldern wie der Internetökonomie treffen zu können, werden Geschäftsmodelle zur Analyse des Tätigkeitsfeldes und der Handlungsoptionen und zur Bewertung der mit ihnen einhergehenden ökonomischen Konsequenzen benötigt.[21] Das Web 2.0 bringt dabei neue Anforderungen an die Adäquatheit von Geschäftsmodellen.[22] Geschäftsmodelle sollen vor allem den Ordnungsrahmen bieten, unternehmerische Wertschöpfungsaktivitäten strukturiert zu analysieren.[23]

Der Definition TIMMERS folgend und analog zu HOPPE und BREITNER können entsprechend des in Abbildung 2 dargestellten Ordnungsrahmens als relevante Teilmodelle Marktmodell, Aktivitätenmodell und Kapitalmodell identifiziert werden.[24] Im Aktivitätenmodell werden die für die Befriedigung der Kundenbedürfnisse notwendigen Aktivitäten beschrieben. Die Analyse der dar-

---

[19] Die erste Bank in Second Life wurde (mangels Aufsicht) schon zahlungsunfähig. Vgl. http://www.spiegel.de/netzwelt/spielzeug/0,1518,499620,00.html [09-09-09]. Das Verbot von Glückspielen hat die Ginko-„Bank" in die Zahlungsunfähigkeit geführt, nachdem ein Blogger auf die bestehenden Ungereimtheiten und falschen Zahlen hingewiesen hat. Die virtuellen Einlagen von derzeit 200.000 bis 700.000 $ stehen momentan zur Disposition. Zum Schneeballsystem und der fehlenden Bankenaufsicht der Kommentar eines Bloggers, vgl. http://www.bernd-schmitz.net/blog/index.php/2007/08/23/die-ginko-financial-bank-stellt-alle-konten-ein/.

[20] Im diesem Punkt ist die „Gesellschaft" schon teilweise aktiv geworden; deutsche Anwender haben schon ein „Arbeitsamt" gegründet. Vgl. Casati et al. (2007). Dabei ist zu beachten, dass diese Arbeitsvermittlung auch der Regelung bedarf: Verwerfungen im Arbeitsmarkt zu Beginn der Industrialisierung, also einem gesellschaftlichen Status, der vielleicht mit dem derzeitigen Entwicklungsstandard von Second Life zu vergleichen ist, wurden von privaten Arbeitsvermittlern ausgenutzt und teilweise verursacht.

[21] Vgl. Grob, vom Brocke (2006), Porter (2001), Stähler (2001), 48 f. sowie der Überblick bei Rentmeister, Klein (2003).

[22] Vgl. Bohl, Manouchehri, Winand (2007), S. 28.

[23] Vgl. Magretta (2002).

[24] Vgl. Timmers (1998), Hoppe, Breitner (2003).

aus resultierenden Kosten und Erlöse erfolgt im Kapitalmodell. Im Marktmodell werden die Wettbewerbsbedingungen und Besonderheiten auf dem zugrunde liegenden Markt beschrieben – für die hier im Folgenden beschriebenen Geschäftsmodelle ist das Second Life.

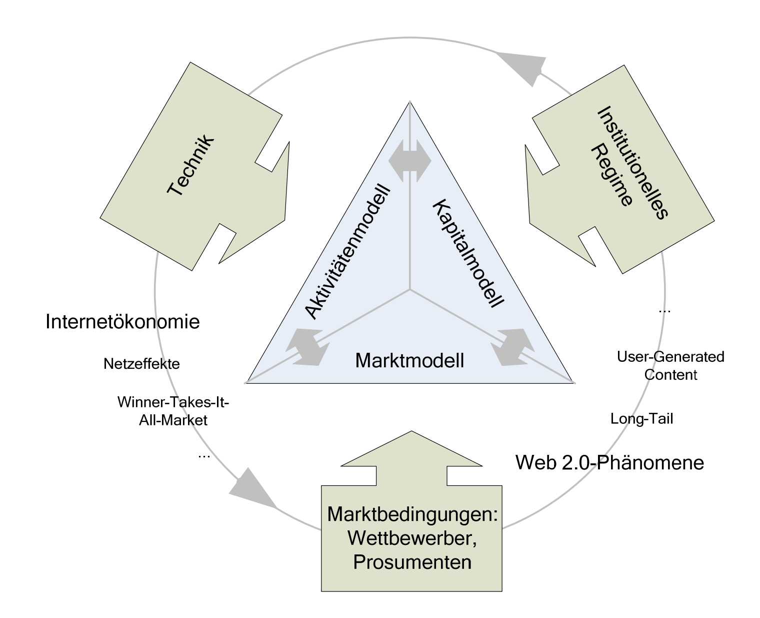


Abbildung 2: Ordnungsrahmen für Geschäftsmodelle im Web 2.0[25]

Dabei sind die Besonderheiten der Internetökonomie und des Web 2.0 zu berücksichtigen, vor allem das Vorhandensein von Netzeffekten und den daraus resultierenden „Winner-Takes-It-All-Market". Daneben sind im Kontext von Web 2.0 mit User-Generated Content und dem so genannten „Long-Tail" Aspekte dieser Marktentwicklung nicht zu vergessen.[26] Wesentliche Gestaltungsparameter und Erfolgstreiber sind die technische Ausgestaltung, die Gestaltung des institutionellen Regimes und die Behandlung der spezifischen Wettbewerbskräfte, die im Prosumermarkt des Web 2.0 entstehen.

4.1.2 Klassifizierung Geschäftsmodelle und Second Life

Bei der Untersuchung von real ausgeübten und potenziellen Geschäftsmodellen im Kontext von Second Life wird deutlich, dass nicht ein oder wenige Archetypen von Geschäftsmodellen klassifizierbar sind, da Second Life ein deutlich weiteres Spektrum als andere Plattformen (mit Ausnahme des Internets selber) bietet. Dabei geht es nicht nur um Prozesse, die erst durch Second Life ermöglicht werden, sondern auch um solche, die durch Second Life verändert werden können. Zur Klassifizierung der unterschiedlichen Modelle von Geschäftsmodellen bietet sich im Zuge der Aktivitätenanalyse eine Klassifikation nach dem Ort der Leistungserstellung und dem überwiegenden Ort der Aktivitäten an.[27]

---

25 Erweiterung des Ansatzes von HOPPE und BREITNER für das Web 2.0, vgl. Hoppe, Breitner (2003), S. 4.

26 Vgl. Alby (2007), Vossen, Hagemann (2007).

27 Einem Ansatz zur Gliederung nach MALONE ET AL. wird hier nicht gefolgt, da die Aussagen einerseits zu generisch sind und damit keine weitergehenden Aussagen erlauben. Andererseits wurde dieses Modell auch noch nicht angepasst an die spezifische Situation von Second Life. So ist die Produktion oder der Handel von virtuellen Menschen, i. e. Avataren, dort z. B. erlaubt. Vgl. Malone et al. (2006). Ebenso wird hier auch nicht der Ansatz von PORTER verfolgt, der zwischen kommerziellen und nicht kommerziellem Ziel und Hintergrund unterscheidet, vgl. Porter (2004). Weiterhin nicht die Topologie von XYZ, weil hier explizit ökonomische Sachverhalte betrachtet werden.

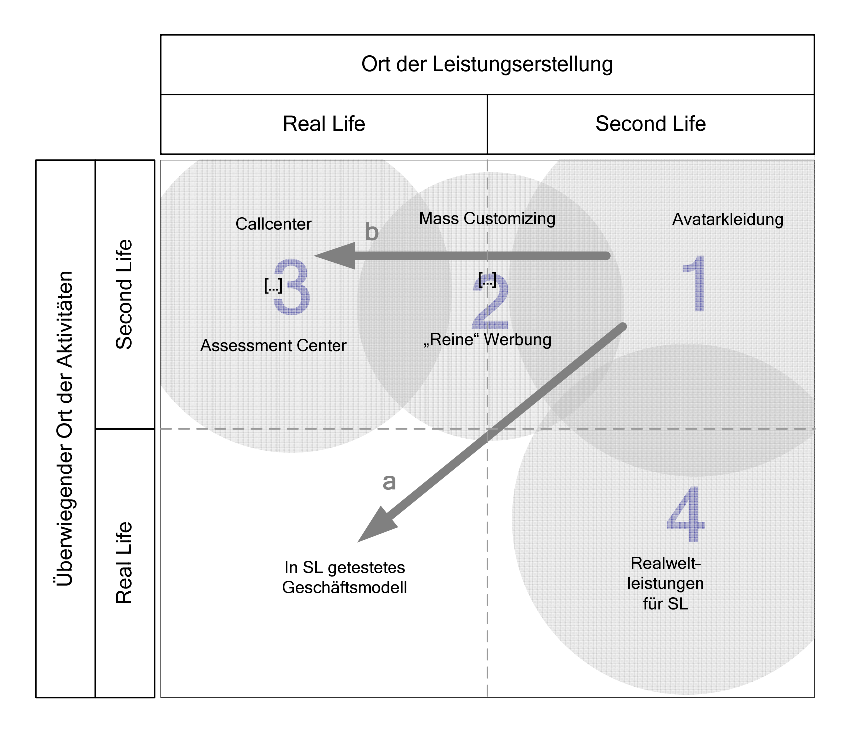


Abbildung 3: Ordnungsrahmen für Geschäftsmodelle mit Second Life

Das Kontinuum dieser unterschiedlichen Möglichkeiten ist in Abbildung 3 dargestellt. Im Fokus stehen Geschäftsmodelle, die innerhalb von Second Life stattfinden (1), neben solchen, bei denen Aktivitäten in Second Life einer Leistungserbringung im realen Leben dienen. Diese können in eher kundennahe (Werbung, Marketing etc.) (2) gegenüber Unterstützungsfunktionen (HR, Call-Center, E-Learning etc.) (3) unterschieden werden. Daneben sind Leistungen aus der Realwelt für Second Life vorstellbar (4). Weiterhin sind auch Verlängerungen innerhalb der Matrix denkbar: Produkte oder ganze Geschäftsmodelle können in Second Life getestet entwickelt werden, bevor diese in das reale Leben überführt werden (a). Auch im Bereich der Unterstützungsleistungen ist ein Übergang denkbar: Ein Zeitarbeitsdienstleister könnte Mitarbeiter seiner Second Life-Dependance im realen Leben übernehmen (b).

4.2 Geschäftsmodelle mit Second Life

Innerhalb von Second Life ist ein variantenreiches Spektrum von verschiedenen Leistungen möglich. Erwartungsgemäß sind diese Geschäftsmodelle unter den ersten, die erprobt wurden. Eine Übersicht der im Folgenden behandelten Klassen von Modellen gibt Abbildung 4.

Die häufigsten Aktivitäten in Second Life betreffen Community-Modelle, öffentliche Rollenspiele und Repräsentationsaufgaben analog zum klassischen WWW. Wenn dieses von nicht-kommerziellen Interessensgruppen durchgeführt wird, ist oft nicht klar, ob hier nur Marketingabsichten dahinter stecken oder der Ansatz, die eigene Organisation in einer virtuellen Welt des Web 2.0 zu etablieren. Second Life zeichnet als echten Web 2.0-Ansatz aus, dass von der Community erstellte hochwertige Strukturen, wie Gebäude und ganze Spiele, existieren. Genauso ist es aber auch Web 2.0-typisch, dass hier die Ertragsmöglichkeiten für Dritte beschränkt sind, da

Anwender in der Regel nicht für selbsterstellte Inhalte bezahlen.[28] Aus der Sicht der Anwender als Prosumenten ist das Potenzial wegen der geringen Kosten recht hoch.

Second Life-inhärent	
	allgemeine Community und Repräsentationsaufgaben aus dem WWW
	Second Life bezogene Dienstleistungen/Güterproduktion
	E-Learning-Angebote über Second Life
	virtuelle Produkte für das „echte Leben“
	nichtselbstständige Arbeit in Second Life
Kundenbezogen	
	Kommunikationsunterstützung, Werbung
	Unterstützung realer Produkte
	Mass Customization
	Open Innovation
	Marktforschung
Indirekte Leistungsbereiche	
	Human Resources: Recruiting, Assessment Center etc.
	Human Resources: Corporate E-Learning
	inter-/intraorganisationale Koordination, Wissensmanagement, Kollaboration
	Virtual Sourcing: Callcenter, Fernarbeitsplätze etc.
	R&D-Labore, Sozialforschung
für Second Life	
	Consulting
	Coding
	Construction

Abbildung 4: Übersicht Geschäftsmodelle im Kontext von Second Life

Interessanter ist die kommerzielle Produktion von Second Life-bezogenen Dienstleistungen und virtuellen Gütern. Dies ist in den folgenden Bereichen möglich: Kleidungen, Avatare und Avatarteile, Einrichtungen, Events, Partys, Reisebüro, Werbung, Medien, Informationsbroker, Trading, Bank, Land, Glücksspiel und sonstige Dienstleistungen, wie Heiratsgestaltung oder körpernahe Dienstleistungen. Die Erfolgschancen sind bei diesen avatarnahen Leistungen eher gering.[29] Erfolgversprechende Modelle sind daneben – wie in der realen Welt – Intermediärmodelle wie Banken und Makler, bei denen Transaktionskosten gesenkt werden oder Informationsvorteile ausgenutzt werden. Wirklich profitable Geschäftsmodelle sind am ehesten zu erwarten, wenn systematische Probleme von Second Life mit dem Geschäftsmodell behoben werden.[30] Problematisch an all diesen Modellen ist, dass sie enorm abhängig von der Entwicklung des zugrunde liegenden Marktes, also dem Erfolg des Geschäftsmodells Second Life, sind. Zudem sind die zivil- und strafrechtlichen Möglichkeiten bisher zu eingeschränkt, die eigenen Interessen in

---

28 Zu der Wahrnehmbarkeit und Sinnhaftigkeit von Werbung in SL siehe auch Bell (2007), Placentra, Gerteis (2007).

29 Es gibt einige Ausnahmen, die als „Stars der Branche“ vergleichbar mit den Top-Modeschöpfern des realen Lebens sind.

30 Vgl. Seidler (2007).

Konfliktfällen durchzusetzen. Dabei besteht keinerlei Sicherheit, da wegen fehlender Institutionen Geschäftsmodelle kurzfristig verboten, straflos kopiert oder durch technische Änderungen obsolet werden können. Ebenso besteht die Gefahr, dass die User entsprechende Leistungen als Prosumer selber erbringen. Wenn mit virtuellen Produkten im realen Leben Profit gemacht werden soll, muss die Darstellung in 3D einen deutlichen Mehrwert gegenüber klassischer Internetdarstellung erbringen.[31] Echtes Erlöspotenzial entfaltet sich hier vor allem, wenn Anwendungen gefunden werden, bei denen – wie im Beispiel geschildert – in der Verbindung von Internet und dreidimensionaler Darstellung und der Nutzung von Community oder Web 2.0-Phänomen ein echter Mehrwert liegt. Eine Dienstleistung, für die dieses zutrifft, sind dedizierte E-Learning-Angebote auf Second Life.[32]

Dabei ist es entweder möglich, für zuvor beschriebene inhärente Leistungen Arbeit in Second Life anzunehmen oder dafür – und für alle im Wesentlichen in Second Life stattfindenden – Arbeiten über Second Life-Mitarbeiter zu verpflichten. Angestellte können auch für Second Life-übergreifende Aktivitäten eingesetzt werden (siehe unten). Für Arbeitgeber kann dieses als potenziell ertragreich erscheinen, da geringe Transaktionskosten durch die Ortsunabhängigkeit entstehen. Zudem ist es Arbeitgebern möglich, Kohorten günstigster Arbeiter für Informationsarbeiten und kundennahe Dienstleistungen ohne räumliche Probleme aus Schwellen und Entwicklungsländern zu rekrutieren.[33] Residents aus den unterentwickelten Ländern können über Second Life jedoch an einem weltweiten Arbeitsmarkt für Informationsarbeit partizipieren. Damit wird der Bogen zu Geschäftsmodellen gespannt, bei denen Aktivitäten auf Second Life mit denen in der „realen Welt“ verbunden werden.

Der bisher häufigste Ansatz im Bereich Marketing ist die Kommunikation über Second Life, ein „Just to be there“-Konzept zur Repräsentation, dessen Mehrwert zweifelhaft ist.[34] Ebenso wie für SL-inhärente Leistungen gilt hier, dass Second Life eine besondere Rolle spielen wird, wenn eine dreidimensionale Darstellung oder community- bzw. avatarbasierte Kommunikationsformen einen Mehrwert darstellen. Diese Mechanismen können zur Erhöhung der Interaktivität mit dem Kunden genutzt werden. Dann können reale Produkte mit einem Mehrwert wie erweiterten After Sales Services oder interaktiven Bedienungsanleitungen über Second Life versehen werden. Die dreidimensionale Darstellung von Produkten und die Erhöhung der Interaktivität ermöglichen auch den Einsatz von Konzepten wie der Mass Customization oder Open Innovation.[35]

Die intra- und interorganisationelle Koordination von Wertschöpfungsketten und die Auslagerung von Unterstützungsleistungen in Second Life ist ein weiteres Tätigkeitsfeld. Aufgrund des erhöhten und stärkeren persönlichen Interaktionspotenzials innerhalb von Second Life sind

---

[31] Ein Beispiel hierfür findet sich im geplanten Geschäftsmodell eines Fitnessstudios, das im Kontext des Businessplan-Wettbewerbes von McKinsey entwickelt wurde, vgl. Seidler (2007).

[32] Dabei erscheint Second Life als Infrastruktur auch noch kosteneffizienter als andere (z. B. eigene) Plattformen zu sein, insbesondere da die Entwicklungskosten auf Second Life vergleichsweise niedrig sind und durch Userbeteiligung zudem noch gesenkt werden können; vgl. Breuer (2007). Besonders hohes Potenzial ist in der handlungsorientierten Ausbildung zu sehen, die sonst in Fallstudien oder Plan- und Rollenspielen personalintensiv an physisch einem Ort umgesetzt werden müssen.

[33] FRIEDMAN beschreibt diesen Effekt von Offshoring und Mass Collabration als Verflachung der Welt – etwas, das auch in Second Life durch die Darstellung des Grids anschaulich visualisiert wird. Vgl. Friedman (2006).

[34] Wahrnehmbarkeit und Akzeptanz von solchen Werbeauftritten in Second Life sind nach ersten Untersuchungen sehr gering, vgl. Bell (2007).

[35] Vgl. Meier, Piller (2001), S. 14 ff.

Auslagerungen von Prozessen im Bereich des Wissensmanagements, der Interaktion zwischen Kooperationspartnern und von bestimmten Back- und Frontoffice-Prozessen im virtuellen Raum unabhängig von der realen räumlichen Verteilung vorstellbar.[36] Naheliegend ist daher auch die Eignung von Second Life gegenüber klassischen Internetanwendungen im Bereich der Human Ressources. Neben reiner Kontaktierung und Vermittlung von Arbeitskräften, die auch über das WWW erfolgreich praktiziert werden, könnten auf Second Life wegen der hohen Personalisierung und Interaktionsfähigkeit Recruitmentprozesse über Wettbewerbe, Werbung, Probearbeiten und Assessmentcenter unterstützt werden.

Letztendlich sind auch klassische Tätigkeiten wie Consulting, Coding und Construction mit einer Spezialisierung für Second Life ein – wie einzelne Beispiele belegen[37] – potenziell ertragreiches Geschäft. Die Grenzen zu rein Second Life-basierten Modellen sind dabei fließend, je nachdem, wo der Schwerpunkt der Tätigkeiten liegt. Im Wesentlichen liegt der Erfolg der beschriebenen Geschäftsmodelle an der Entwicklung des zugrunde liegenden Marktes, also von Second Life, und der Erweiterung der technischen Möglichkeiten.

4.3 Second Life als Geschäftsmodell

Im Folgenden wird das Geschäftsmodell Second Life untersucht, da von dessen Entwicklung das Gelingen der beschriebenen Geschäftsmodelle abhängt. Daher ist die Beschaffenheit und Gestaltung der Wettbewerbsstruktur für Second Life von vitaler Bedeutung und soll im Folgenden angelehnt an eine Branchenstrukturanalyse nach PORTER untersucht werden, die in Abbildung 5 dargestellt ist.[38] Dabei ist zu antizipieren, dass im Web 2.0 in der Regel nicht mehr überwiegend klassische Lieferanten-Kunden-Beziehungen existieren, sondern vor allem Prosumentenbeziehungen. Daneben ist dem Konzept von Second Life immanent, dass auch finanzielle Anreize für die Produktion von Content gesetzt werden und nicht nur wie sonst üblich nur meritoökonomische.[39] Daher werden in Second Life potenziell Prosumenten attraktiert, die über ein anderes, auch finanziell orientiertes Anreizsystem verfügen. In dem Geschäftsmodell von Second Life müssen differenziertere Nutzerpräferenzen und ein komplexeres technisches und funktionales Betätigungsfeld beachtet werden, anders als bei typischen Web 2.0-Communities üblich. Dabei kann bei einem Anbieter von kostenpflichtigen Diensten im Rahmen des Web 2.0 eine Unterscheidung in Abnehmer und Lieferanten nicht vorgenommen werden. Vielmehr sind die Kunden hier auch Produzenten ihres eigenen Konsums, also so genannte Prosumenten.[40] In Second Life lassen die auf dieser Ebene beteiligten Akteure sich in solche mit potenziell kommerzieller Absicht (*Professionals*) und tendenziell nur konsumierende Kunden (*Privats*) klassifizieren. Diese Rollenverteilung ist nicht disjunkt – einzelne Akteure wechseln zwischen diesen Rollen oder nehmen diese parallel war.

---

36 Beispielsweise im so genannten agilen Information Management, vgl. Bauer, Mandl (2007).

37 Vgl. Casati et al. (2007) oder im Web z. B. http://www.seminal3d.de/html/seminal3d/A---Mainmenu/a---Ueberblick/03---Umsetzung/1---CrossMediaAnsatz.html [09-09-09] oder http://secondlife.bokowsky.net/de/produkte/online_communities/ [09-09-09].

38 Zu PORTER allgemein vgl. Kutz (2006), Wettbewerbsanalyse und die Implikationen des Internets, vgl. Porter (2001).

39 Zu diesem und weiteren Motivatoren vgl. Stoll (2006), Untersuchungen am Beispiel des Videoportals youtube.com haben gezeigt, dass die Hauptmotivationen der Anwender, eigenen Content zu veröffentlichen, meritoökonomischer Natur war, vgl. McKinsey (2007), S. 1.

40 Vgl. Tapscott, Williams (2006).

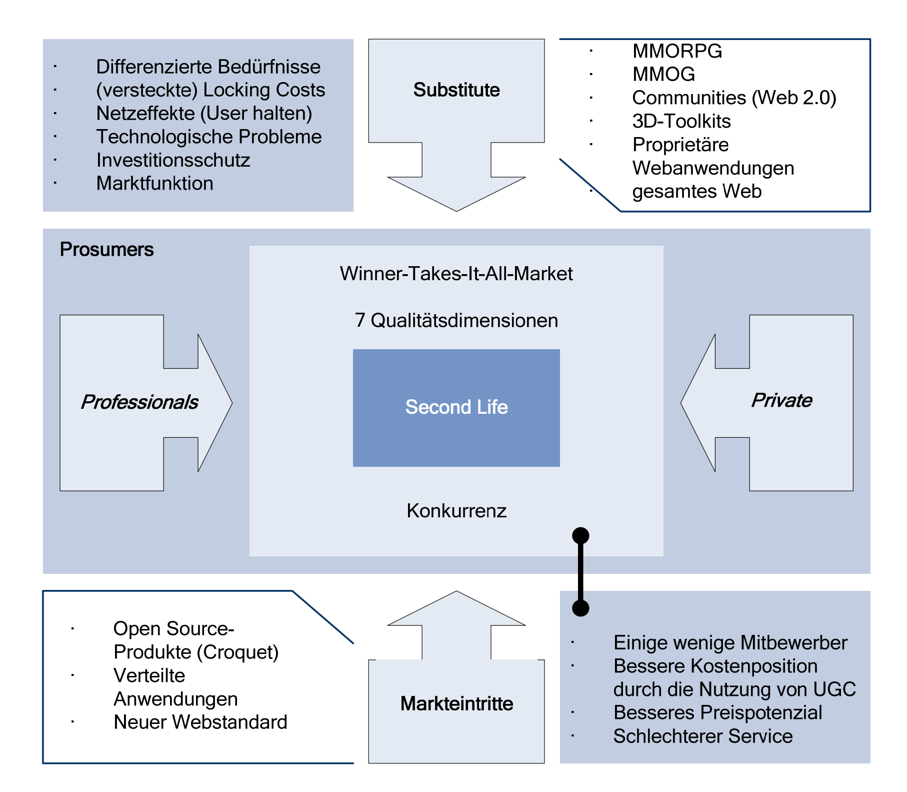


Abbildung 5: Wettbewerbskräfte beim Geschäftsmodell virtuelle Welt am Beispiel von Second Life

Die Betrachtung dieser Beziehungen ist im Kontext von Web 2.0 im Allgemeinen und im Speziellen bei der Betrachtung der derzeitigen Marktsituation von Second Life besonders wichtig, weil hier die wichtigen Netzwerkeffekte realisiert werden, die einen Erfolg in einem Winner-Takes-It-All-Market ausmachen können.[41] Dies ist im Web 2.0 bei Prosumentenbeziehungen von besonderer Relevanz, weil die Erstellung von Inhalten durch Anwender die wachsenden Rückkopplungsschleifen der Netzeffekte positiv verstärkt. Second Life muss sich daher schnell als beherrschende Plattform in einem evivierenden Markt etablieren, wofür die Kundenzufriedenheit mit den für Web 2.0-Aspekte angepassten Qualitätsanforderungen an Portale eine wesentliche Rolle spielt. Als wesentliche Erfolgsfaktoren für die Portalqualität können identifiziert werden: Content (Inhalte), Challenge (Herausforderung), Communication (Kommunikation), Commerce (Transaktionen), Configuration (Gestaltung/Konfiguration) und Customer Care (Unterstützung und Betreuung).[42] Im Kontext virtueller Welten sollte zudem die institutionelle und technische Ausgestaltung dieser Plattform einbezogen werden – insbesondere, wie diese Arrangements die Transaktionskosten und hier vor allem die Suchkosten beeinflussen (Context).

---

41 Vgl. Vossen, Hagemann (2007), O'Reilly (2005), Alby (2007).

42 Vgl. Bauer, Hammerschmidt (2004).

Qualitätsdimension	Privates	Professionals
Content	User-Generated Content, macht Attraktivität aus, ist selbst einfach steuerbar	Herausforderung: Nutzung von und Konkurrenz zu User-Generated Content
Challenge	User-Generated Content, macht Attraktivität aus, ist selbst einfach steuerbar	Herausforderung: Nutzung von und Konkurrenz zu User-Generated Content
Communication	Wichtig, macht Attraktivität aus, bisher recht gute Möglichkeiten	Möglichkeit zur Massenkommunikation fehlt
Commerce	Käuferschutz reicht wegen geringer Beträge von Privates in der Regel aus	Kein Investitionsschutz, muss nachgebessert werden
Configuration	nicht optimaler Client – Misserfolgsfaktor – per Open Source evtl. ausgeräumt	Wesentliche Anbindungsparameter fehlen
Customer Care	Seitens Second Life für die Anwender eher unwichtig (Netzeffekte haben auf die Plattformauswahl von Privates höhere Auswirkung)	Sehr wichtig, Nachbesserung notwendig
Context	Sechs Verhaltensregeln reichen in der Regel	Wesentlich mehr notwendig

Abbildung 6: Ausprägungen der Qualitätsdimensionen in Abhängigkeit von der Rolle

Die Gestaltung der Wechselbarrieren ist aus der Sicht von Second Life als gelungen anzusehen. Durch einen für 3D-Internetplattformen niederschwelligen Zugang werden geringe Locking Costs signalisiert – die Partizipation an Second Life und die Erstellung von nicht portierbaren Inhalten verursacht schleichend hohe spezifische Investitionen, die bei den Privates psychologische Kosten bei einem Systemwechsel verursachen würden. Die verglichen mit Konkurrenzplattformen niedrigen Entwicklungskosten könnten auch Professionals zu spezifischen Investitionen veranlassen. Beiden Gruppen ist gemein, dass sie einen Wechsel auf eine andere Plattform scheuen, da durch die Proprietät der getätigten Entwicklung hohe Ersatzinvestitionen notwendig wären. Die meisten Konkurrenten haben bei weitem nicht die Größe von Second Life erreicht. Für professionelle Anbieter ist das zudem in der Regel dort teuer, weil die Erstellung aufwendiger ist.[43] Andererseits sind hier teilweise die technischen Möglichkeiten überlegen und die Unternehmen betreiben mehr Customer Care für ihre Premiumkunden, sodass diese auch mehr Vertrauen gewinnen. Direkte Substitute sind alle möglichen Anwendungen, die im klassischen Internet gleiche Funktionalitäten anbieten. Second Life hat hingegen den Vorteil der Integration in einer Plattform. Daneben stellen vor allem MMORGs Substitute dar.[44] Daneben ist vorstellbar, eine virtuelle Welt für eigene Produkte zu entwickeln.[45] Eintrittsbarrieren existieren in diesem Markt nur technische – weder Second Life noch ein anderer Wettbewerber hat bisher eine kritische Größe erreicht, die einen erfolgreichen Eintritt eines neuen Wettbewerbers ausschließt. Daher könnte vor allem ein übergreifender Standard gefährlich werden, der das etablierte, bewährte und elaborierte Web um 3D Funktionalität erweitert.[46]

---

43 Vgl. Breuer (2007).

44 Vgl. Breuer (2007).

45 Vgl. http://www.spiegel.de/netzwelt/spielzeug/0,1518,479848,00.html [09-09-09].

46 IBM arbeitet z. B. daran und soll Gerüchten aus Branchenkreisen zufolge über 3.000 Leute auf das Projekt gezogen haben.

Zusammenfassend besitzt Second Life – gerade als größter Anbieter in einem Netzeffekt basierten Markt – das Potenzial, weiter enorm zu wachsen und das Internet zu verändern. Das verlangt eine Handhabung des Netzeffektes. Viel bedeutender erscheint eine Anpassung des Institutionellen Regimes an die gewachsenen Erfordernisse. Das notwendige schnelle Wachstum kann Second Life nicht aus eigener Kraft stemmen. Der Weg, Inhalte von den Anwendern erzeugen zu lassen, sollte auch auf die technische Infrastruktur übertragen werden. Die Offenlegung des Quellcodes des Clienten ist ein Schritt in die richtige Richtung. Andererseits sollte Linden Lab versuchen, strategische Allianzen mit großen Unternehmen und Organisationen der realen Welt einzugehen, die über spezifische Kenntnisse in wichtigen Bereichen verfügen, insbesondere des Designs von institutionellen Arrangements und der Transaktionsabwicklung. Beispiele wären hier neben Regierungsorganisationen z. B. große Banken oder Finanzhandelsplätze. Dabei darf allerdings nicht die Beteiligung interessierter Anwendergruppen aus dem Anwenderkreis von Second Life vernachlässigt werden.

5 Fazit

Ziel des Beitrags war eine Aufarbeitung der virtuellen Welt „Second Life“ aus wirtschaftlicher Perspektive. Dabei wurde deutlich, dass die wirtschaftlich-soziale Entwicklung, die die technische Infrastruktur erlaubt, so derzeit nicht tragbar ist. Auf Basis der gewonnenen Erkenntnisse wurden Geschäftsmodelle im Kontext von Second Life untersucht. Dazu wurde zunächst ein Ordnungsrahmen für Geschäftsmodelle im Web 2.0 eingeführt. Anhand dieses Ordnungsrahmens wurden reale und potenzielle unternehmerische Tätigkeitsfelder hinsichtlich der verbundenen Aktivitäten- und Kapitalmodelle untersucht. Als potenzielle Wirkungsbereiche wurden Second Life-inhärente Modelle, Modelle, die einen Übergang zwischen Realwelt und Second Life bieten und Übertragungsstrategien untersucht. Dabei konnte festgestellt werden, dass viele Modelle, die nur auf Second Life fokussieren, zwar grundsätzlich profitabel, aber nicht überdurchschnittlich wachstumsstark sind und somit nur der persönlichen Existenzsicherung dienen können. Dies könnte ein Erwerbsmodell für Residents aus Schwellenländern darstellen. In Second Life hingegen sind dieselben Geschäftsmodelle besonders aussichtsreich wie in der realen Welt: Insbesondere solche, die auf Basisinnovationen basieren oder klassische Intermediärmodelle darstellen, haben überdurchschnittliche Wachstumsaussichten. Daneben wurde festgestellt, dass Modelle, die die Wertschöpfung der realen mit der zweiten Welt verbinden, vielversprechend sind, vor allem, wenn aus der dreidimensionalen Darstellung Mehrwert generiert werden kann oder die geringen Transaktionskosten z. B. durch die Anwendung als Offshoring-Plattform ausgenutzt werden. Die Erfolgsaussichten der meisten dieser Modelle hängen von der weiteren technischen Entwicklung und der zukünftigen Stellung von Second Life im Markt ab. Bei der Untersuchung von Second Life fiel auf, dass Web 2.0-spezifisch nicht Kunden-Lieferanten-Beziehungen, sondern Provider-Prosumer-Beziehungen vorliegen. Im Gegensatz zu der sonstigen Praxis herrschen auf Second Life allerdings nicht meritoökonomische Anreizsysteme für die Erzeugung von User-Generated Content, sondern es existiert ein reales finanzielles Anreizsystem. Damit nehmen die Prosumenten zwei unterschiedliche (nicht-disjunkte) Rollen wahr, die des *Privates* und des *Professionals*. Für die weitere Entwicklung von Second Life wurde die Partizipation von Professionals als vitaler Faktor herausgestellt. Um die Marktattraktivität für diese Zielgruppe zu erhöhen, sind einige Maßnahmen notwendig. So ist z. B. zur Sicherung der spezifischen Investitionen der Teilnehmer eine Fortentwicklung des institutionellen Regimes unabdingbar. Entwicklungs- und Forschungsbedarf bleibt vor allem in den Fragen der weiteren

Entwicklung von Second Life. So bestechend das Konzept ist, dass das Team von Linden Lab um den Gründer *Rosedale* präsentiert, so kritisch ist die Berichterstattung inzwischen geworden. Ebenso sind nicht alle – insbesondere die letzten – Statistiken besonders vielversprechend. Diese kritischen Stimmen können allerdings nicht darüber hinwegtäuschen, dass das Konzept einer Plattform, dass die Gestaltungsprinzipien des Web 2.0 auf die Spitze treibt, indem es den Nutzern erlaubt, auch Zweck und Art ihrer Interaktion selber zu gestalten, sehr vielversprechend ist. Nach einer Untersuchung des Beratungsunternehmens *Fittkau und Maaß* ist der Anteil der aktiven Second Life-Nutzer in Deutschland mit Abstand am höchsten in der Gruppe der Teenager.[47] Zumindest letzteres ist ein Fakt, der durchaus für Second Life spricht.

Literaturverzeichnis

Alby, T. (2007), Web 2.0 – Konzepte, Anwendungen, Technologien, 2. Aufl., München 2007.

Bauer, H. H., Hammerschmidt, H. (2004), Kundenzufriedenheit und Kundenbindung bei Internet-Portalen – Eine kausalanalytische Studie in: Bauer, H. H., Rösger, J., Neumann, M. (Hrsg.): Konsumentenverhalten im Internet, München 2004, S. 189-214.

Bauer, N., Mandl, P. (2007), Agiles Informationsmanagement – Informationsbereitstellung in Unternehmen mit Web 2.0, in: HDM – Praxis der Wirtschaftsinformatik, 2007, H. 255, S. 88-96.

Bell, M. (2007), Ladenschluss in Second Life, in: W&V, 2007, H. 23, S. 52-53.

Bohl, O., Manouchehri, S., Winand, U. (2007), Unternehmerische Wertschöpfung im Web 2.0, in: HDM – Praxis der Wirtschaftsinformatik, H. 255, 2007, S. 27-36.

Bonstein, J. (2006), „Was du machst, ist ganz egal“, in: Der Spiegel 39/2006, S. 116-118.

Breuer, M. (2007), Second Life und Business in virtuellen Welten, White Paper Berlin 2007, S. 1-62, im WWW unter http://www.pixelpark.com/de/pixelpark/_ressourcen/attachments/publikationen/0703_White_Paper_Second_Life_e7_Pixelpark.pdf [11-09-07].

Casati, R., Matussek, M., Oehmke, P., Uslar, M. v. (2007), Alice im Wunderland, in: Der Spiegel, 2007, H. 8, S. 150-163.

Engelmann, S. (2007), Dem virtuellen Rausch folgt der reale Kater, in: JIM – Journal of Interactive Media, 2007, H. 1, S. 14-17, zugl. im WWW unter http://www.uni-hohenheim.de/komm-in/uploads/pics/JIM_1.pdf [09-09-09].

Friedman, T. L. (2006), Die Welt ist flach – Eine kurze Geschichte des 21. Jahrhunderts, Frankfurt a. M. 2006.

Grob, H. L., vom Brocke, J., Bensberg, F. (2005), Finanzwirtschaftliche Bewertung von Geschäftsmodellen im E-Learning: Konzeption, Methoden und Perspektiven, in: Breitner, M. H., Hoppe, G. (Hrsg.), E-Learning – Einsatzkonzepte und Geschäftsmodelle, Heidelberg 2005, S. 101-116.

Grob, H. L., vom Brocke, J. (2006): Internetökonomie, Das Internet im Fokus hybrider Systeme, in: Grob, H. L., vom Brocke, J. (Hrsg.), Internetökonomie, Ein interdisziplinärer Beitrag zur Erklärung und Gestaltung hybrider Systeme, München 2006, S. 1-20.

Hillenbrand, T. (2006), 300 Dollar fürs virtuelle Wohnzimmer, in: Financial Times Deutschland, FTD.de, 28.06.2006, im WWW unter http://www.ftd.de/technik/it_telekommunikation/90138.html [11-09-07].

[47] Vgl. o. V. (2007).

Hopf, K., Braml, B. (2007), Virtuelle Kinderpornographie vor dem Hintergrund des Online-Spiels Second Life, in: ZUM, 2007, H. 5, S. 354-363.

Krasemann, H. (2006), Onlinespielrecht – Spielwiese für Juristen, in: MMR, 2006, H. 6, S. 351-357.

Kutz, O. (2006), Strategische Analysetechniken, in: Zerres, C., Zerres, M. P. (Hrsg.), Handbuch, Marketing-Controlling, 3. Aufl., Berlin 2006, S. 11-45.

Linden Lab (2007), Linden Lab Introduces Identity Verification to Second Life, Pressemitteilung vom 29. August 2007, San Francisco 2007, im WWW unter http://lindenlab.com/press/releases/08_29_07 [09-09-09].

Lischkar, K. (2007a), Flickr verbietet Deutschen Nacktfotos, im WWW unter http://www.spiegel.de/netzwelt/web/0,1518,488542,00.html [09-09-09].

Lischkar, K. (2007b), Flickr filtert den Protest, im WWW unter http://www.spiegel.de/netzwelt/web/0,1518,489273,00.html [09-09-09].

Lischkar, K. (2007c), Gnadenlose Richter gefährden Web 2.0 in Deutschland, im WWW unter http://www.spiegel.de/netzwelt/web/0,1518,490006,00.html [09-09-09].

Magretta, J. (2002), Why Business Models Matter, in: MBR, 2002, H. 5, S. 3-8.

Malone, T. W., Weill, P., Lai, R. K., D'Urso, V. T., Herman, V. T., Apel, T. G., Woerner, S. L. (2006), Do Some Business Models Perform Better than Others?, MIT Sloan Working Paper 4615-06, 2006, im WWW unter http://ssrn.com/abstract=920667 [09-09-09].

McKinsey (2007), How companies can make the most of user-generated content, in McKinsey Quarterly, August 2007, im WWW unter http://www.mckinseyquarterly.com/article_abstract.aspx?ar=2041&l2=17&l3=104&srid=17.

Meier, R., Piller, F. T. (2001), Strategien zur effizienten Individualisierung von Dienstleistungen, in: Industrie-Management, 17. Jg., 2001, H. 2, S. 13-17.

Niemeier, J. (1998), Internet-Communities als Geschäftsmodell – Von der Web-Präsenz über den elektronischen Marktplatz zur virtuellen Gemeinschaft, in: zfo, 1998, H. 4, S. 220-223.

o.V. (2007), Second Life – die aktiven Nutzer, Studie Fittkau & Maaß, Juni 2007, im WWW unter http://www.w3b.org/second_life.html [09-09-09].

O'Reilly, T. (2007), What Is Web 2.0 – Design Patterns and Business Models for the Next Generation of Software, im WWW unter http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html [09-09-09].

Pham, K. (2007), StudiVZ-Mitglieder treiben Extremisten in die Flucht, im WWW unter http://www.spiegel.de/netzwelt/web/0,1518,503637,00.html [09-09-09].

Placentra, G., Gerteis, T. (2007), Second Life – Umsatzmotor oder Seifenblase, in: W&V, S. 42-44.

Porter M. E. (2001), Strategy and the Internet, in: Harvard Business Review, 49. Jg., 2001, H. 3, S. 63-78.

Porter, C. A. (2004). A Typology of Virtual Communities: A Multi-Disciplinary Foundation for Future Research, in: Journal of Computer-Mediated Communication, 10. Jg., 2004, H. 1, im WWW unter http://jcmc.indiana.edu/vol10/issue1/porter.html [09-09-09].

Rentmeister J., Klein S. (2003) Geschäftsmodelle. Ein Modellbegriff auf der Wagschale, in: ZfB-Ergänzungsheft, o. Jg., 2003, H. 1, S 17-30.

Rippert, S., Weimer, K. (2007), Rechtsbeziehungen in der virtuellen Welt, in: ZUM, 2007, H. 4, S. 272-281.

Seidler, C. (2007), Auf der Suche nach den „Second Babes", im WWW unter http://www.spiegel.de/netzwelt/spielzeug/0,1518,500776,00.html [20-08-07].

Smith, D. A., Key, A., Raab, A., Reed, D. P. (2003), Croquet – A Collaboration System Architecture, working paper, im WWW unter http://www.croquetconsortium.org/images/2/2b/2003_Croquet_Collab_Arch.pdf [09-09-09].

Stähler P. (2001), Geschäftsmodelle in der digitalen Ökonomie, Köln 2001.

Stöcker, C. (2007), Sex-Stalker im Studentennetz, im WWW unter http://www.spiegel.de/netzwelt/web/0,1518,450866,00.html [09-09-09].

Stoll, B. L. (2006), Spass und Software-Entwicklung. Zur Motivation von Open Source Programmieren, Univ.-Diss. Zürich 2006, im WWW unter www.dissertationen.unizh.ch/2006/luthigerstoll/diss.pdf [05-06-07].

Tapscott, D., Williams, A. (2006), Wikinomics – How Mass Collaboration Changes Everything, New York 2006.

Timmers, P. (1998), Business Models for Electronic Markets, in: Electronic Markets, 8. Jg., 1998, H. 2, S 3-8.

Vossen, G., Hagemann, S. (2007), Unleashing Web 2.0 – From Concepts to Creativity, Amsterdam 2007.

Zimprich, S. (2006), Inflationsangst in der Parallelwelt, in: Financial Times Deutschland, FTD.de, 1.11.2006, im WWW unter http://www.ftd.de/technik/medien_internet/127075.html [11-09-07].

Sicherheit und Vertrauen im Wandel vom Read zum Read/WriteWeb

Gunnar Thies

1 Einführung

Wer das Phänomen des Web 2.0 beschreiben möchte, verwendet meist Schlagwörter wie *User-Generated Content*, also von Nutzern erstellter Inhalt, und *Communities*, soziale Netzwerke, die sich aus Internetnutzern aus aller Welt zusammensetzen und täglich Veränderungen unterliegen. Ebenfalls in diesem Kontext werden *Mash-Ups* genannt: Dies sind Anwendungen, die versuchen, bereits vorhandene Anwendungen miteinander nutzbringend zu verbinden oder mit Mehrwert für die *Community* anzureichern. Es ist erstaunlich, wie Web 2.0-Anwendungen wie Pilze aus dem Boden schießen und immer wieder neue Ideen verwirklicht werden oder alte Ideen durch *Mash-Ups* in neuem Licht erscheinen. Die Vielfalt der Anwendungen reicht hierbei von Freizeitcharakter über Kommunikationsplattformen bis hin zu kommerziellen Zwecken.

Man darf jedoch nicht vergessen, dass das Internet ein global geöffnetes Medium ist, in dem die unterschiedlichsten Interessen der Nutzer aufeinander treffen. Das man aus diesem Grund nicht jedem Inhalt einer Anwendung, einer Mail oder einer Webseite des World Wide Webs blind vertrauen darf, mussten und müssen viele Nutzer immer wieder feststellen: Als Beispiele seien hier Spam-E-Mails, *Phishing* und *Code Injection* aufgeführt. Diese Methoden werden genutzt, um beispielsweise Bankdaten der Nutzer auszuspionieren oder an Passwörter für Onlineanwendungen zu gelangen. Um eine Web 2.0-Anwendung zu benutzen, werden im Normalfall Benutzerprofile durch den Anbieter angelegt, bei denen der Nutzer sich mit einer E-Mail-Adresse registriert. Dadurch wird er Teil einer Onlinegemeinschaft. Gedanken um die Sicherheit des Dienstes im Hinblick auf die Verwendung der persönlichen (Profil-)Daten machen sich wohl die wenigsten Personen, wenn man den Aufschwung von Web 2.0-Anwendungen betrachtet. Es scheint ein implizites Vertrauen zum Anbieter zu bestehen. In diesem Zusammenhang sagt Tantek Celik, *Chief Technologist* bei *Technorati* in [TW07]:

> *„Something really interesting happens when you trust your customers, they trust you".*

Diese Aussage beschreibt das Verhältnis von Kunde zu Anbieter. Zusätzlich dazu wird aber das Vertrauen zwischen einzelnen Nutzern im Web 2.0 immer wichtiger, denn viele *Communities* und Dienste würden ohne ein grundlegendes Vertrauen unter den Mitgliedern nicht funktionieren.

Im Folgenden wird daher zunächst der Wandel des Vertrauensbegriffs des Web 2.0 im Gegensatz zum Web 1.0 skizziert, und wie man als Anbieter das Vertrauen zwischen Kunde und sich und den Kunden bzw. Nutzern untereinander stärken kann. Anschließend wird auf den Sicherheitsbegriff in der Informationstechnologie im Web 2.0-Kontext eingegangen und der enge Zusammenhang zwischen Vertrauen und Sicherheit bei Web 2.0-Anwendungen erläutert. Außerdem wird *SingleSignOn* als eine Technik zur Steigerung der Sicherheit und des Komforts für

Web 2.0-Nutzer vorgestellt. Dann folgt noch eine kurze Betrachtung von Beispielen, in denen Konzerne oder sogar Regierungen das Vertrauen von Web 2.0-Communities missbraucht haben. Eine Zusammenfassung und ein Ausblick bilden den Abschluss.

2 Web 2.0: Vom Read zum Read/WriteWeb

Der Name *Read/WriteWeb* tauchte bereits im Jahre 2003 auf, als Richard MacManus[1] seine Arbeiten am heute hoch angesehenen Read/WriteWeb-Blog aufnahm. Der Gedanke dahinter ist, dass das World Wide Web (WWW) nicht mehr nur rein lesend benutzt werden soll, sondern vielmehr die Inhalte von den Nutzern gelesen **und** geschrieben werden können. Natürlich war das WWW von Anfang an so konzipiert, dass Benutzer Inhalte beispielsweise per Webformular beisteuern konnten, aber erst mit der großen Verfügbarkeit des Internets, vor allem durch günstige Breitbandanschlüsse, kam der Web 2.0-Boom in Schwung. Dion Hinchcliffe, Web 2.0-Veteran und Gründer von Hinchcliffe & Company, erläutert hierzu in [1]:

„Thus, Read-write Web + People Using It = Web 2.0.“

Das heutige WWW wird seiner Meinung nach erst durch die Verwendung durch die Nutzer zum Web 2.0. Menschen benötigen zur Nutzung eines Dienstes oder einer Seite aber ein gewisses Vertrauen in die Technik und den Anbieter. In diesem Zusammenhang soll Vertrauen anhand einer für den Web 2.0-Bereich sehr passenden Definition aus der Soziologie von Niklas Luhmann kurz definiert werden: Vertrauen ist demnach ein „Mechanismus zur Reduktion sozialer Komplexität“. So wird auf Erfahrung und Intuition basierend meist „die positive Entwicklung von Ereignissen“ (hier sinngemäß der Wahrheitsgehalt einer Information) angenommen, da eine rationale Überprüfung meist nicht möglich ist.[2]

Der Fokus im Web 1.0 lag hierbei hauptsächlich auf dem Vertrauensverhältnis zwischen Kunde und Anbieter z. B. beim elektronischen Einkaufen (E-Commerce, vgl. [AGW+05], S. 44 ff.). Eine Unternehmung kann hierbei hauptsächlich durch Informationen über das Unternehmen selbst Vertrauen aufbauen (vgl. [2]). Dazu bietet man:

- Daten zur Organisation (und über die Mitarbeiter),
- eine klar verständliche und gesetzeskonforme Datenschutzerklärung,
- eine durch vertrauenswürdige Dritte (TÜV, VeriSign) bestätigte Professionalität und
- ein professionelles Design der Webseite mit regelmäßigen Updates und wenigen Fehlern.

Das Vertrauen von Kunden in den Zahlungsverkehr beim E-Commerce wird teilweise auch dadurch genährt, dass Zahlungen zwischen Geschäftspartnern über eine dritte Partei, z. B. *PayPal*[3] getätigt werden (vgl. [VH07], Kapitel 1). Dabei werden die Bank- oder Kreditkartendaten der Nutzer nur dort gespeichert und ein E-Commerce-Anbieter wie *Amazon* oder *eBay* rechnet über *PayPal* ab. Der Vorteil ist darin zu sehen, dass die Geschäftspartner keine Bankdaten

---

1 MacManus arbeitete im SiliconValley und als Manager für Topfirmen in Neuseeland.

2 http://de.wikipedia.org/wiki/Vertrauen.

3 https://www.paypal.com/de/.

austauschen müssen, dennoch muss wenigstens das Vertrauen zur Zahlungstransaktion ausführenden dritten Partei gegeben sein.

Auch große IT-Unternehmen wollen ihr Vertrauensverhältnis zum Kunden verbessern bzw. erhalten und tragen dem Boom des Web 2.0 bereits Rechnung: Es gibt dort Regelungen zum *Bloggen*[4] (so genannte *Blog Policies*), wie beispielsweise von der Firma DELL die „Online Communication Policy“ (vgl. [3]). Weitere Unternehmen wie IBM oder Groove gehen sogar so weit, das *Bloggen* während der Arbeitszeit zu erlauben (vgl. [4]). Dies gibt den Nutzern das Gefühl, mehr hinter die Kulissen blicken zu können, da sie mit Firmenmitarbeitern direkt kommunizieren und deren *Blogs* kommentieren können, anstatt mit einer gesichtslosen Firma vorlieb nehmen zu müssen.

2.1 Vertrauen

Das Web 2.0 ändert nicht nur die Sichtweise von Unternehmen, sondern auch die der Nutzer, denn in vielen Fällen findet ein Wechsel der „Geschäftspartner“ statt: Der Inhalt einer Seite wird nun zu großen Teilen von den Nutzern erstellt und nicht mehr von einer verantwortlichen Firma. Technik tritt hierbei nun mehr in den Hintergrund und der Mensch als Anbieter bzw. Redakteur von Informationen wird wichtig. Wie aber kann man einem Menschen vertrauen, den man nicht kennt? Nötig werden hier Instrumente, um andere Nutzer einschätzen zu können: bei dem Auktionshaus *eBay* wird das Vertrauen beispielsweise durch öffentliche Bewertungsprofile gestärkt.[5] Bei der Fülle von Web 2.0-Anbietern aus den verschiedensten Bereichen wie Freizeit und Vergnügen (z. B. *MySpace*, *YouTube*, *Secondlife*), Information und Wissen (z. B. *Wikipedia*, *Digg*) und E-Commerce (z. B. *eBay*, *Amazon*) wandelt sich die Frage „Kann ich dem Anbieter vertrauen?“ immer mehr zu „Kann ich den Benutzern vertrauen?“.

2.1.1 Nutzergeneriertes Vertrauen

Das häufigste Werkzeug, um Vertrauen zwischen den Nutzern eines Web 2.0-Dienstes aufzubauen, ist das Erstellen von Bewertungsprofilen durch die Nutzer selbst. Hierfür gibt es den Ansatz, dass ein Nutzer für jeden generierten Inhalt von anderen bewertet werden darf, was man als *User-Generated Trust* bezeichnet. Bei *eBay* wird jeweils für eine erfolgte Auktion eine Bewertung für Käufer und Verkäufer ermöglicht. Somit kann vor der Abgabe eines Auktionsgebots nachgeschaut werden, ob der Verkäufer vertrauenswürdig ist, indem man seine positiven Bewertungen überprüft. Bei *YouTube* oder *Clipfish* werden die vom Nutzer hochgeladenen Videos durch andere Nutzer auf einer Skala von 1 bis 5 bewertet. Der Nachteil dieser Bewertungssysteme ist, dass jeder Web 2.0-Dienst ein eigenes hat und damit jeder Nutzer in verschiedenen Systemen verschiedene Bewertungsprofile besitzt.

Ein diesen Nachteil behebender Ansatz wird durch den Dienst *Venyo*™ angeboten: Hier wird ein Benutzerprofil angelegt, das Bewertungen aus verschiedensten Web 2.0-Anwendungen heraus ermöglicht. Dafür wird innerhalb eines beliebigen Dienstes ein Link angezeigt, der es

---

4 *Bloggen* bedeutet soviel wie: schreiben in ein Onlinetagebuch, ein so genanntes *Blog*.

5 Diese Bewertungsprofile können aber durch Falscheinträge oder teilweises Löschen der Bewertungen „geschönt“ werden.

erlaubt, eine Bewertung für den Nutzer zu verfassen, der bei *Venyo* gespeichert wird. So können beispielsweise Photos bei *Flickr*, Artikel bei *Blogger* und Videos bei *YouTube* bewertet werden, wobei alle Bewertungen im so genannten *Vindex* bei *Venyo* zusammengefasst werden. Natürlich muss für eine erfolgreiche Verwendung von *Venyo* ein Großteil der Web 2.0-Dienste das Bewerten in *Venyo* erlauben, was bisher noch nicht der Fall ist (vgl. [5]). Daher ist *Venyo* im Grunde ein weiteres soziales Netzwerk, mit dem Unterschied, dass es über mehrere Web 2.0-*Communities* verteilt funktionieren soll.

Aber auch ohne einen Dienst wie *Venyo* funktioniert das Grundprinzip sozialer Netzwerke bereits durch die grundlegenden menschlichen Verhaltensweisen erstaunlich gut: Inhalte, die einen Nutzer stören oder nicht interessieren, werden meist ignoriert oder sogar geächtet. Tun dies viele Nutzer, werden automatisch Strukturen erreicht, bei denen unliebsame Nutzer des Netzwerks unbeachtet bleiben und so automatisch an den Rand des Netzwerks verbannt werden. Implizit wirkt in einem sozialen Netzwerk auch eine Art des aus der Kryptologie bekannten Prinzips des *Web-of-Trust*[6]: „Ich vertraue jedem, dem jemand vertraut, dem ich vertraue." Daher gibt es auch in fast jeder Web 2.0-Anwendung eine Art „Freundesliste", in die Nutzer eingetragen werden, denen man vertraut.

2.1.2 Anbietergeneriertes Vertrauen

Bewertungsprofile alleine schaffen noch kein umfassendes Vertrauen, da diese in einigen Fällen leicht manipulierbar sind (im Falle von *eBay* ist dies schon des Öfteren geschehen). Daher muss auch der Dienstanbieter zur Vertrauensbildung beitragen. *provider generated trust* basiert im Web 2.0 zum Großteil auf der Authentifizierung des Nutzers: Es sollte sicher sein, dass der Nutzer auch der ist, der er vorgibt zu sein. Dies ist bei Diensten wie *YouTube* oder *Flickr* weniger relevant als bei E-Commerce-Diensten wie *eBay*, sollte aber dennoch nicht vernachlässigt werden.

Es gibt hierbei verschiedene Schritte zur Vertrauensbildung: Der erste geschieht bereits bei der Registrierung am Dienst, der zweite danach in der Verwendung des Dienstes und dem Transparenzgrad der Nutzerprofile (vgl. [2]).

Registrierung

Die Sicherung der Identität einer Person kann verschieden stark ausgeprägt sein: die einfachste Möglichkeit ist das Anmelden am Dienst per E-Mail-Adresse und dem Beantworten einer Aktivierungs-E-Mail des Anbieters. Somit ist lediglich die Existenz der E-Mail-Adresse überprüfbar. Wird zusätzlich ein Aktivierungscode verlangt (beispielsweise per *CAPTCHA*[7]), kann zumindest das automatisierte Registrieren durch *Bots*[8] erschwert werden, die Identität ist dadurch aber auch nicht bestätigt. Die sicherste (und noch praktikable) Methode ist immer noch das An-

---

6 http://de.wikipedia.org/wiki/Web_of_Trust.

7 ***C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part*. Dabei werden z. B. einfache Buchstabenfolgen mittels Bildfilter verzerrt. Menschliche Nutzer können dies relativ leicht lesen, Computer haben dagegen ihre Schwierigkeiten.

8 Computerprogramme, die automatisch Aufgaben ausführen, wie z. B. das Anmelden an einem Dienst.

schreiben per Post an eine Heim- oder Geschäftsadresse, die gerade im E-Commerce die Identität sicherzustellen versucht.[9]

Verwendung des Dienstes

Um während der Verwendung des Dienstes das Vertrauen in den Dienst und seine Nutzer zu steigern, können zunächst die Benutzerprofile innerhalb der *Community* öffentlich (in verschiedenen Detailgraden) angezeigt werden – natürlich unter Beachtung des Datenschutzgesetzes. Eine zusätzliche Maßnahme ist das Kenntlichmachen von identifizierten Nutzern: so kann man beispielsweise per E-Mail-Registrierung den Dienst nutzen, aber erst durch eine postalische Registrierung „aufsteigen" zum identifizierten Nutzer. Dies wird beispielsweise bei *Amazon.de* durch den Zusatz *Real Name* bei der Anzeige von Nutzernamen verdeutlicht. Es wird in [2] auch die Möglichkeit der Echtzeitkonversation (z. B. mittels *Skype*[10]) vorgeschlagen, was das Vertrauen natürlich steigern kann, da man mit einer realen Person spricht.

Man darf auch nicht unerwähnt lassen, dass auf der anderen Seite die E-Commerce-Anbieter immer das Risiko eingehen, dass es einen Nutzer nicht gibt, oder dessen Daten falsch sind. Das Vertrauen, dass z. B. *Amazon* einem Nutzer bei einer Bestellung per Nachnahme an eine Privatadresse entgegenbringt, ist nicht zu vernachlässigen.

2.2 Sicherheitsaspekte

Der Begriff Sicherheit muss in diesem Kontext zunächst in zwei Teilbereiche gegliedert werden: die Datensicherheit und die Computersicherheit (Programmierung, Datenübermittlung). Der Begriff Datensicherheit stammt aus dem Datenschutz und ist juristischer Natur: Damit ist die Behandlung der anfallenden Profil- und Nutzungsdaten gemäß des Datenschutzgesetzes (vgl. [6]) adressiert. Die Computersicherheit soll den Schutz vor Ausfall und Manipulation gewährleisten sowie den unerlaubten Zugriff auf Daten von Nutzern verhindern (vgl. [7]).

2.2.1 Datensicherheit

Die Sicherheit der Daten liegt in der Hand des jeweiligen Anbieters und hängt natürlich auch stark von der Art der Anwendung und dem rechtlichen Standort der Unternehmung ab. Da viele Web 2.0-Dienstleister in den USA. residieren, greift das deutsche Datenschutzgesetz nicht. Die Diskrepanz zwischen dem US-amerikanischen Verständnis von Datenschutz und dem Deutschlands (bzw. Europas) tritt z. B. bei den Bestimmungen zu Datensammlung und -zusammenführung zutage: In den USA ist dies für private Firmen ohne Einschränkungen erlaubt, im Gegensatz dazu ist im deutschen Datenschutzrecht eine enge Zweckbindung für das Sammeln von Daten nötig und die Zusammenführung von Daten ist ebenfalls stark reglementiert (vgl. [8]). Dabei sollte man nicht nur die wenigen – relativ unkritischen – Profildaten im Auge behalten, die im Zweifel nur die E-Mail-Adresse und einen Benutzernamen nebst Passwort beinhalten, sondern ebenso alle Verlaufsdaten, Kontakte und andere persönliche Daten, wie z. B. geschriebe-

---

[9] Eine einwandfreie Identifizierung ist nur durch das persönliche Erscheinen und Vorzeigen eines Ausweisdokumentes machbar, wie es bei der Vergabe von Zertifikaten und Kreditkarten getan wird. Dies ist aber kein gangbarer Weg für Web 2.0-Dienste.

[10] *Skype* ermöglicht IP-Telefonie zwischen zwei oder mehreren Internetnutzern.

ne E-Mails. Die aktuelle Gesetzeslage in Deutschland mit der Anpassung des so genannten Hackerparagrafen (§ 202 StGB) und der damit geschürten Diskussion über die Onlinedurchsuchungen und den „Bundestrojaner" zeigen aber, dass auch die Datenschutzgesetze in Deutschland nicht unantastbar sind (vgl. [9] und [10]).

2.2.2 Sicherheit der Technik

Die technische Umsetzung von Web 2.0-Diensten basiert in der Regel auf einer in einem Browser aufrufbaren Webseite, die durch die Verwendung von *Ajax* (einer *JavaScript*-Technologie) benutzerfreundlich und desktopähnlich zu bedienen ist (vgl. [VH07], Kapitel 3.1). *JavaScript* war noch vor einigen Jahren als „sicherheitslückenbehaftet" verpönt, ist nun aber – quasi über Nacht – salonfähig geworden: obwohl die damals genannten Probleme wie *CrossSite-Scripting* und *Code Injection* nach wie vor existieren (vgl. dazu [11]). Die meisten Web 2.0-Anwendungen würden ohne *JavaScript* bzw. *Ajax* aber nicht oder nur umständlich funktionieren. Bei der Diskussion um technische Sicherheit können in diesem Zusammenhang drei Aspekte identifiziert werden: die Programmsicherheit, die sichere Datenübertragung und der sichere Identitätsnachweis.

Sicherheit eines Programms

Die Sicherheit (Fehlerfreiheit) eines Programms wird gerade dann wichtig, wenn viele Nutzer den Dienst verwenden und relevante Daten abgelegt sind. Zum einen wirkt sich eine Sicherheitslücke auf das Vertrauen zum Anbieter aus, zum anderen können vertrauliche Daten ausgespäht werden. Das größte Problem sei *Code Injection* mittels *Ajax* bzw. *JavaScript,* womit die Codequalität nicht gerade erhöht wird (vgl. [BMS07]).

Sicherheit bei der Datenübertragung

Eine sichere Datenübertragung ist mittels des *Secure Sockets Layer (SSL)*-Verfahrens möglich. Hierbei verläuft die gesamte Datenübermittlung zwischen Browser des Nutzers und Server verschlüsselt. Dazu wird zwischen zwei Teilnehmern ein geheimer (Sitzungs)Schlüssel ausgetauscht, der dann für jede weitere ausgetauschte Nachricht als Basis für die Verschlüsselung verwendet wird (vgl. [12]). Diese Methode wird mittlerweile als Standardverschlüsselung für Onlinebanking flächendeckend eingesetzt.

Sicherheit der Identität

Web 2.0-Dienste benutzen in verschiedenem Umfang Benutzerprofile und daher Logins, um einen Nutzer zu identifizieren. In den meisten Fällen ist dies bei jeder Anwendung ein eigenes Authentifizierungssystem. Eine Möglichkeit, dies einfacher und vertrauenswürdiger zu gestalten, ist die Verwendung eines *SingleSignOn* (SSO)-Systems, das einen einmaligen Login ermöglicht, und dann auf jeder Seite, die dieses System unterstützt, quasi automatisches Einloggen ermöglicht. Ein großer Vorteil ist hierbei, dass der Nutzer sich nur ein einziges Passwort für viele Anwendungen merken muss und dass er in jedem System mit derselben Identität bzw. mit demselben Login Zugang hat (je nach Implementierung). Zwei bekannte Systeme sind hierbei *OpenID* (vgl. [13]) und Microsofts *CardSpace* (vgl. [14]). Im Bereich der Universitätsbibliotheken in Deutschland ist mittlerweile auch *Shibboleth* (vgl. [15]) als eine SSO-Lösung akzeptiert. Im

Folgenden wird das im Web 2.0-Bereich führende System *OpenID* exemplarisch für die Funktionsweise von SSO erläutert. Wichtig hierbei ist, dass diese Methoden keine Mechanismen zur direkten Vertrauensbildung sind (*trust mechanisms*), sondern lediglich SSO realisieren. Die Überprüfung der Identität eines Nutzers muss bei der Registrierung geschehen.

SingleSignOn mittels OpenID

SingleSignOn besteht im Wesentlichen immer aus einem oder mehreren *Service-Providern* (SP) und mindestens einem *Identity-Provider* (IP). Der IP verwaltet die Benutzerdaten, gibt auf Anfrage Daten des Benutzers heraus und bestätigt die Identität des Nutzers gegenüber einem SP. Ein SP bietet eine Dienstleistung – hier einen Web 2.0-Dienst – an und authentifiziert seine Benutzer, indem ein IP angefragt wird. Abbildung 1 (Quelle: www.macprofessionell.de) zeigt eine grobe Übersicht über die erforderlichen Schritte.

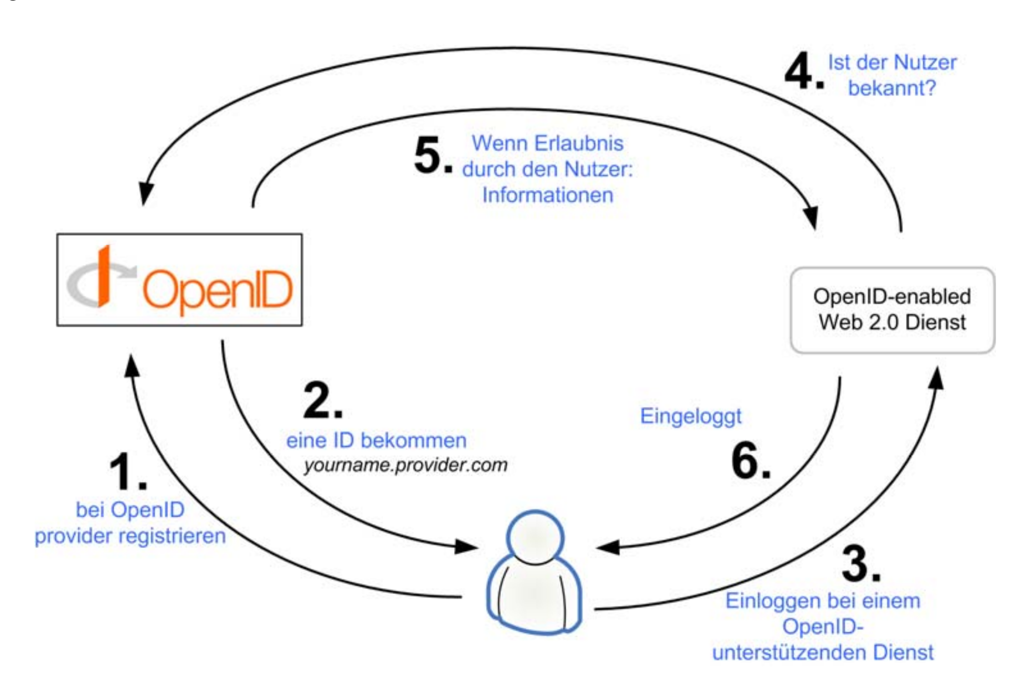


Abbildung 1: Ablauf des SingleSignOn per *OpenID*

Im ersten Schritt registriert sich der Nutzer bei einem *OpenID*-Anbieter (z. B. *myopenid.com* oder *verisignlabs.com*) und erhält von diesem im zweiten Schritt einen Aliasnamen und Passwort (hier: *yourname.provider.com*). Im dritten Schritt loggt man sich mit diesem Aliasnamen bei einem Dienst – also dem SP – ein, der *OpenID* unterstützt. Solche Seiten sind beispielsweise unter [16] zu finden. Der Dienst fragt dann bei dem *OpenID*-Anbieter – dem IP – anhand des Aliasnamen im 4. Schritt an, ob der Nutzer eingeloggt ist. Kann der IP dies positiv beantworten, bekommt der SP die für ihn vom Nutzer freigegebenen Daten mitgeteilt und erkennt den Nutzer als eingeloggt. Ist der Nutzer noch nicht eingeloggt, wird er zur IP-Webseite weitergeleitet und dort aufgefordert, sich am IP mit Benutzernamen und Passwort anzumelden. Anschließend wird er wieder zur Seite des SP weitergereicht. Nach diesem Schritt ist der Nutzer am IP und am Dienst eingeloggt. Die Funktionalität des SSO greift nun jedes Mal, wenn sich der Nutzer auf weiteren *OpenID*-gestützten Seiten einloggen möchte: Hier muss er lediglich seinen *OpenID*-

Aliasnamen angeben, Schritt 4 bis 6 laufen automatisch im Hintergrund ab und er ist eingeloggt.[11]

Es entstehen dadurch mehrere Vorteile für den Nutzer:

- Der *OpenID*-Account kann mehrere Identitäten verwalten und für jeden Dienst nur bestimmte Daten freigeben (die Kontrolle, auf welche persönliche Daten ein Web 2.0-Dienst zugreifen darf, verbleibt somit beim Nutzer).
- Das redundante Eingeben von Daten bei der Registrierung an einem Dienst (der *OpenID* unterstützt) entfällt, da diese einfach per Mausklick von der Identitätskarte zugeordnet werden können.

Unter [16] sind aktuell 343 Dienste gelistet, die *OpenID* als Loginmechanismus erlauben: darunter bekannte Web 2.0-Dienste, wie beispielsweise *Zooomr*, *digg* und *grou.ps*. Um sich dem Umgang mit *OpenID* zu erleichtern, lässt das Firefox Plugin *SeatBelt* von *VeriSign* installieren, das den aktuellen Loginstatus anzeigt.

Microsofts *CardSpace* kann ebenso zur Nutzung eines SSO-Mechanismus implementiert werden, erlaubt aber darüber hinaus durch die Nutzung des *Webservice-Stacks* und das Mitschicken von Sicherheits-Tokens auch die Anbindung von Anwendungen außerhalb des Browser-Fokus. Dadurch ist *CardSpace* weitaus mächtiger als *OpenID*, wird aber auch wesentlich komplizierter (vgl. [18] und [19]).

2.3 Zusammenhang von Sicherheit und Vertrauen

Die Begriffe Vertrauen und Sicherheit bedingen sich gegenseitig stark. Je sicherer eine Webapplikation erscheint, unabhängig ob Web 2.0, Onlinebanking oder E-Commerce, desto eher sind Nutzer oder Geschäftspartner dazu bereit, der Anwendung Vertrauen entgegenzubringen. Dies spiegelt sich dann im Normalfall auch in der Nutzerzahl wider.

So kann man klar herausstellen, dass Vertrauen durch sinnvoll eingesetzte Technik stark vorangetrieben wird. Zum einen wird durch sichere Technik bei der Registrierung und im Betrieb einer Anwendung das Vertrauen des Nutzers zum **Anbieter** automatisch stärker. Zum anderen kann man als Nutzer, wenn man der Technik und dem Anbieter vertraut, auch dem **Nutzer** des gleichen Dienstes ein hohes Vertrauen entgegenbringen.

3 Implikationen für Business-Anwendungen

Neben den vielen Erfolgsgeschichten von Web 2.0- und E-Commerce-Anwendungen, zu denen sicherlich *Amazon* und *eBay* gehören, gibt es auch die Fälle, in denen Unternehmen das Medium WWW missbräuchlich oder zumindest werbewirksam unglücklich benutzt haben. Im Folgenden werden drei Fälle aus der Praxis vorgestellt, die das zuvor definierte Vertrauen – hier als Wahrheitsgehalt einer Information verstanden – von Nutzern verspielten.

---

[11] Die genaue Spezifikation von OpenID ist unter [17] nachzulesen.

TheHub von Walmart

Walmart versuchte sich im Juli 2006 an einer eigenen *Community* mit dem Namen *TheHub*. Die Anwendung folgte dem Beispiel von *MySpace,* zielte auf das Marktsegment der jugendlichen Walmart-Kunden ab und hatte das Motto „Back to school". Eines der Probleme des Dienstes war die Kommerzialisierung durch viele Werbeeinblendungen von Walmart-Produkten. Ein weiteres – in sozialen Netzwerken viel größeres Problem – war die Glaubwürdigkeit: Denn Walmart stattete Mitarbeiter mit vorgetäuschten Mitgliederaccounts aus, die einem der vier Stereotypen Kaufsüchtiger, Skateboarder, Footballspieler oder R&B-Künstler zugeordnet wurden. Dann wurde mit diesen Accounts über die „coolen" Produkte von Walmart diskutiert. Dies wirkte so unglaubwürdig auf die echten Nutzer, dass man das Projekt aus Mangel an Vertrauen und dem daraufhin entstehenden Nutzerschwund kurzerhand nach nur zehn Wochen wieder einstellte (vgl. [20]). Hieran ist zu sehen, dass die Sicherheit in erster Linie kein Problem war, sondern das Vertrauen durch das vorgetäuschte soziale Netzwerk erschüttert wurde.

„Technosexual" von Calvin Klein (CK)

Eine Frankfurter Werbeagentur wollte den Begriff „technosexual" als Markenname eines neuen Duftes von Calvin Klein innerhalb der deutschen Web 2.0-*Community* etablieren. Dazu wurde eine *Blog*-Telenovela mit mehreren Akteuren inszeniert, die über einen gewissen Zeitraum in verschiedenen *Blogs* und Foren immer wieder den Begriff „technosexual" einwarfen. Dieser Werbetrick flog auf, da einige Nutzer die verschiedenen *Blogs* verfolgten und die teilweise doch ungewöhnlich offenen Beschreibungen einer Dreiecksbeziehung und immer wieder der Nennung des zu bewerbenden Begriffes einfach zu unglaubwürdig fanden. Schließlich gab die Werbeagentur zu, dies alles nur als Werbekampagne für Calvin Klein aufgesetzt zu haben (vgl. [21] und [22]). Ob dies nun ein Imageverlust für CK mit sich gebracht hat, oder die Werbewirksamkeit in diesem Fall sogar förderlich war, ist nicht mit Gewissheit bestimmbar. Fakt ist jedoch, dass wiederum die Glaubwürdigkeit innerhalb eines sozialen Netzwerkes (hier in *Blogs*) schwand und damit das Vertrauen in CK missbraucht wurde.

Änderungen an Wikipedia

Virgil Griffith (ein Informatikstudent aus Kalifornien) zeigt mit einem kleinen Projekt – dem *WikiScanner* – eindrucksvoll, dass Firmen und Institutionen Einträge bei *Wikipedia* (zu ihren Gunsten) abändern. So konnte er anhand von Änderungsprotokollen und der zugehörigen IP-Adresse in *Wikipedia* nachweisen, dass beispielsweise Microsoft Artikel über deren *XBox* und andere Produkte abänderte, oder die US-Regierung den *Wikipedia*-Eintrag über das Gefangenenlager Guantanamo schönte. Wiederum ein Fall, in dem das Vertrauen der Nutzer in die *Wikipedia*-Community zunächst etwas gestört war. Die Entwicklung des *WikiScanners* kann aber eben dieses Vertrauen stärken, denn mit diesem Werkzeug können solche eindeutigen Änderungen nachverfolgt werden und der Wahrheitsgehalt einer Information (Vertrauen) wird in gewisser Weise kontrollierbar oder zumindest bewertbar (vgl. [23] und [24]).

Dies sind nur einige kurze Beispiele für die Verwendung und den Missbrauch des Phänomens Web 2.0 für Imagepflege und Marketingzwecke. Aber gerade Unternehmen, deren Geschäftsfeld Web 2.0-Anwendungen sind, sind darauf angewiesen, dieses Vertrauen nicht zu verspielen. Was diese drei Beispiele auch zeigen: Die Sicherheit (vor allem die subjektiv einschätzbare Sicherheit der Programmierung und der Datenübertragung) scheint im Web 2.0-Bereich kein

Problem zu sein. Die meisten Probleme entstehen, wenn das soziale Netzwerk Mängel in der Glaubwürdigkeit erfährt.

4 Zusammenfassung und Ausblick

Vertrauen zu einem Nutzer des Internets oder zu einem Anbieter aufzubauen, ist nicht leicht, da man im Grunde nicht genau weiß, mit wem man es zu tun hat. Gerade in modernen Web 2.0-Anwendungen, die auf nutzergeneriertem Inhalt basieren, muss das Vertrauen implizit vorhanden sein. Man kann dies aber stärken, indem man durch organisatorische Maßnahmen, angefangen bei der Registrierung und Überprüfung einer Identität bis hin zur möglichst großzügigen Informationspolitik, Transparenz schafft. Darüber hinaus lassen sich technische Mittel wie *SingleSignOn*-Mechanismen und *SSL*-Verschlüsselung nutzen, um möglichst nutzerfreundlich eine sichere Plattform anzubieten.

Natürlich ist der Einsatz von sicheren Technologien und möglichst fehlerfreier Programmierung kein Garant für Vertrauen. Sicherheitsspezifische Komponenten werden in der Softwarebranche immer weiterentwickelt: Aber jede neue Technologie muss sich erst bewähren oder wird mehr oder weniger schnell ausgehebelt. In der moderneren Geschichte sei hierfür beispielsweise die WLAN-Verschlüsselung WEP genannt, die mittlerweile innerhalb von Minuten geknackt werden kann. Auch neuere Anwendungen sind davon betroffen, Microsofts Betriebssystem Windows Vista™ verwendet ein komplexes Lizenzsystem, um nur zertifizierte Treiber und Geräte zu betreiben. Dieser Schutz wurde innerhalb weniger Wochen von findigen Hackern ausgeschaltet. Daher muss sich zeigen, ob die Einführung von SSO-Mechanismen wie *OpenID* und *CardSpace* tragfähig sind und sich *SSL* als Standard für Datenübermittlung flächendeckend im Web 2.0-Bereich durchsetzt. Wobei die Frage im Raum steht, ob die Daten, die in Web 2.0-Anwendungen anfallen, einen so schützenswerten Rang einnehmen.

Das Prinzip der sozialen Netzwerke und dem nutzergenerierten Inhalt scheint – unter der Voraussetzung einer originellen Idee – in den meisten Fällen zu funktionieren. Sobald aber die Glaubwürdigkeit einer *Community* infrage gestellt wird oder die Sicherheit der Anwendung nicht gegeben ist, kann einer Web 2.0-Anwendung sehr schnell das Ende drohen.

Literaturverzeichnis

[AGW+05] Alpar, P., Grob, H. L., Weimann, P., Winter, R., Anwendungsorientierte Wirtschaftsinformatik – Strategische Planung, Entwicklung und Nutzung von Informations- und Kommunikationssystemen, 4., überarb. u. verb. Aufl., Braunschweig, Wiesbaden 2005.

[BMS07] Beck, A., Mörike, M., Sauerburger H., Web 2.0: Konzepte, Technologie, Anwendungen, in: Praxis der Wirtschaftsinformatik, HMD, H. 255, 2007.

[TW07] Tapscott, D., Williams, A. D., Wikinomics, How Mass Collaboration Changes Everything, New York 2007.

[VH07] Vossen, G., Hagemann. S., Unleashing Web 2.0: From Concepts to Creativity, San Francisco, CA 2007.

Internetadressen

[1] http://blogs.zdnet.com/Hinchcliffe/?p=41, *A round of Web 2.0 reductionism*, Dion Hinchcliffe, zuletzt abgerufen am 29.08.2007
[2] http://www.webcredible.co.uk/user-friendly-resources/web-credibility/web20-trust.shtml, *The usability & accessibility specialists*, zuletzt abgerufen am 28.08.2007
[3] http://www.dell.com/content/topics/global.aspx/corp/governance/en/online_comm, *Online Communication Policy von DELL*, zuletzt abgerufen am 29.08.2007
[4] http://www.corporateblogging.info/2005/06/policies-compared-todays-corporate.asp, *Vergleich verschiedener Blog-Policies amerikanischer Firmen*, zuletzt abgerufen am 29.08.2007
[5] http://www.venyo.org*, The web 2.0 Trust Provider,* zuletzt abgerufen am 26.08.2007
[6] http://de.wikipedia.org/wiki/Datenschutz, zuletzt abgerufen am 02.09.2007
[7] http://de.wikipedia.org/wiki/Computersicherheit, zuletzt abgerufen am 02.09.2007
[8] http://web.inf.tu-dresden.de/~lvinfhl4/download/dsds/5_nationaler_vergleich.pdf, zuletzt abgerufen am 02.09.2007
[9] http://www.ccc.de/press/releases/2006/20060925/?language=de, Artikel zum „Hackerparagraphen“, zuletzt abgerufen am 31.08.2007
[10] http://www.heise.de/newsticker/meldung/95247, Artikel zur Verfassungskonformität des „Bundestrojaners“, zuletzt abgerufen am 31.08.2007
[11] http://www.heise.de/security/artikel/84149, *Gesundes Misstrauen: Sicherheit von Webanwendungen*, c’t 26/06, zuletzt abgerufen am 31.08.2007
[12] http://de.wikipedia.org/wiki/Transport_Layer_Security, zuletzt abgerufen am 30.08.2007
[13] http://openid.net/, *OpenID Hauptseite*, zuletzt abgerufen am 30.08.2007
[14] http://cardspace.netfx3.com/, Microsoft CardSpace Startseite, zuletzt abgerufen am 31.08.2007
[15] http://shibboleth.internet2.edu/, Shibboleth Projektseite, zuletzt abgerufen am 31.08.2007
[16] http://openiddirectory.com/, zuletzt abgerufen am 31.08.2007
[17] http://openid.net/specs.bml, *Spezifikation von OpenID*, zuletzt abgerufen am 31.08.2007
[18] http://blog.doubleslash.de/2007/01/25/microsoft-cardspace-so-funktionierts/, zuletzt abgerufen am 31.08.2007
[19] http://msdn2.microsoft.com/en-us/library/aa480189.aspx, zuletzt abgerufen am 31.08.2007
[20] http://www.forbes.com/digitalentertainment/2006/10/02/myspace-walmart-youtube-tech-media-cx_rr_1003walmart.html, zuletzt abgerufen am 31.08.2007
[21] http://www.technosexual.de, *Blogging-Telenovela-Werbeaktion von Calvin Klein*, zuletzt abgerufen am 26.08.2007
[22] http://www.themaastrix.net/?page_id=696, *Blogging-Telenovela-Werbeaktion chronologischer Verlauf,* zuletzt abgerufen am 31.08.2007
[23] http://wikiscanner.virgil.gr/, *Webseite von Virgil Griffith zur Veränderung von Wikipedia durch Firmen und Institutionen*, zuletzt abgerufen am 28.08.2007
[24] http://www.newstube.de/Wissen/Wikipedia-Scanner_sp %FCrt_Manipulationen_auf/, zuletzt abgerufen am 31.08.2007

III Suchen und Sozialisieren

Universelles Suchen im Web – Eine technische, ökonomische und soziale Betrachtung

Sebastian Herwig

1 Von horizontaler über vertikale hin zur universellen Suche

„The library of Alexandria was the first time humanity attempted to bring the sum total of human knowledge together in one place at one time. Our last attempt? Google.“[25]

Brewster Kahle, Gesellschafter und Gründer des Internet Archive.

Wie die Essenz des vorangestellten Zitats andeutet, nehmen Suchmaschinen [28] im Zeitalter des World Wide Webs (kurz: Web) eine immer bedeutendere Stellung ein. Im Hinblick auf die stetig wachsende Flut an Informationsartefakten im Web stellen Suchmaschinen ein nahezu unverzichtbares Instrument für das Auffinden von Informationen sowie zur Navigation und Orientierung in den dynamischen und komplexen Informationsstrukturen des Webs dar. Die Dienste von Suchmaschinen werden dabei zunehmend zum allgegenwärtigen Bestandteil des gesellschaftlichen Lebens und fungieren nicht nur als Einstiegspunkt zum Web, sondern vielmehr zu einem umfassenden Informationsfundus einer Bibliothek der gegenwärtigen Informationsgesellschaft [25].

Aus technischer Sicht betrachtet, finden Suchmaschinen ihren Ursprung im Bereich des *Information Retrieval* [26]. Durch den Aufbau eines Schlagwortindex, basierend auf einer Menge von Dokumenten, ermöglichen es Suchmaschinen, diese Dokumentenbasis mittels Schlagwörtern zu durchsuchen. Das wesentliche Bestreben von Suchmaschinen ist es dabei, Informationen von für den durch eine Suchanfrage dargestellten Informationsbedarf möglichst hoher Relevanz bereitzustellen [26, 27].

Auf diesem Funktionsprinzip aufbauend, sind Suchmaschinen im Web seit Beginn ihres Aufkommens darauf ausgerichtet, einen umfassenden Index aufzubauen, welcher sich möglichst über die gesamte Dokumentenbasis im Web erstreckt. Auf dieser Basis versuchen Suchdienste, eine Suche horizontal über nahezu dem gesamten Themenspektrum des Webs zu ermöglichen [28]. Dieses nunmehr als *horizontale Suche* bekannte Konzept gestattet somit einen breiten Zugriff auf die Informationen im Web. Im Hinblick auf die zunehmende Vielfalt von Informationen bieten horizontale Suchmaschinen jedoch nur begrenzt Zugriffsmöglichkeiten auf themenspezifische Informationen. Als Reaktion darauf haben sich sog. *vertikale Suchmaschinen* [29] herausgebildet, die fokussiert für eine Domäne Suchdienste bereitstellen. Um dies zu realisieren, sind vertikale Suchdienste in ihrer gesamten Funktionalität auf die spezifischen Gegebenheiten der betrachteten Domäne ausgerichtet. Dies äußert sich vor allem darin, dass vertikale Suchdienste im Gegensatz zu horizontalen Suchdiensten auf einem themenspezifischen Schlagwortindex aufbauen. Sie verfügen damit über eine präzisere Kenntnis der semantischen Bedeutungen von Information innerhalb des betrachteten Themenfeldes [30, 14]. Somit erlaubt das Konzept der vertikalen Suche vielmehr eine zielgerichtete Suche, die Informationen von

vermeintlich höherer Relevanz bereitstellt und ermöglicht darüber hinaus auch den Zugriff auf Informationen im so genannten *Deep Web* [35], die über die traditionelle horizontale Websuche nicht erreicht werden können.

Das Aufkommen von vertikalen Suchdiensten lässt die Verdrängung traditioneller horizontaler Suchkonzepte vermuten [30]. Es wird jedoch schnell deutlich, das der Aufbau von vertikalen Suchdiensten vor allem für enggefasste Themenbereiche schon rein aus einer ökonomischen Perspektive oft nicht sinnvoll erscheint [29]. Infolgedessen existieren vertikale Suchdienste derzeit fast ausschließlich für bedeutende Domänen und sind als ergänzendes Konzept zu traditionellen horizontalen Suchdiensten zu sehen. Um jedoch die jeweiligen Vorteile der horizontalen und vertikalen Suche verbinden zu können, bestehen aktuell Bestrebungen, die beiden Suchkonzepte derart zu verknüpfen, dass eine Suche simultan in horizontalen sowie in einer Vielzahl verschiedener vertikaler Suchdienste ermöglicht werden kann und die aus der Suchanfrage resultierenden Informationsartefakte in einer gemeinsamen Ergebnispräsentation aufbereitet werden [30, 14, 19]. Dieses Konzept wird aktuell intensiv unter dem Begriff der *universellen Suche* diskutiert und ist Gegenstand der nachfolgenden Überlegungen. Aufbauend auf der Darstellung der grundlegenden Funktionsweise wird das Konzept der universellen Suche unter ökonomischen und sozialen Gesichtspunkten betrachtet und an zwei Realisierungsbeispielen veranschaulicht. Abschließend wird kurz ein Ausblick auf mögliche Entwicklungstendenzen der Suche im Web gegeben.

2 Universelles Suchen im Web

Mit dem grundlegenden Gedanken, die Vielzahl an thematisch speziellen Informationssilos der vertikalen Suchmaschinen aufzubrechen und diese mit der traditionellen horizontalen Websuche zu verbinden, verfolgt das Konzept der universellen Suche im Wesentlichen einen integrativen Ansatz [14]. Das logische Konzept, welches dem Ansatz der universellen Suche zugrunde liegt, wird bei einer Gegenüberstellung mit dem Konzept aktuell vorherrschender Suchdienste deutlich (vgl. Abbildung 1).

Suchmaschinenbetreiber wie Google, Yahoo! oder Ask tendieren in den letzten Jahren zunehmend zum Aufbau einer Vielzahl spezialisierter Suchmaschinen, wie Bilder-, Video-, News- oder Produktsuchen, ergänzend zum Angebot der traditionell horizontalen Websuche, um den nutzerspezifischen Informationsbedarfen zielgerichteter begegnen und entsprechend relevante Informationen bereitstellen zu können. Dadurch gestaltet sich der Suchprozess bei derzeitigen Suchdiensten sehr diversifiziert, sodass eine Suchanfrage explizit einzeln an verschiedene im Vorfeld ausgewählte Suchmaschinen gestellt werden muss. Dies hat einerseits zur Folge, dass in den Suchprozess häufig nur bekannte Suchmaschinen einbezogen und vermeintlich relevante Suchdienste aufgrund mangelnden Bewusstseins über deren Existenz vernachlässigt werden [14]. Andererseits werden die Ergebnisse der verschiedenen Suchmaschinen in separaten Trefferlisten präsentiert, was die Auswahl relevanter Informationen erschwert und damit den Suchaufwand und die Suchkosten erhöht.

Demgegenüber wird nunmehr durch den integrativen Ansatz der universellen Suche versucht, das Angebot der traditionellen Websuche mit der Vielzahl existierender vertikaler Suchmaschinen derart zu verknüpfen, das eine Suchanfrage simultan an die verschiedenen Suchmaschi-

nen weitergeleitet wird und die jeweils zurückgelieferten Suchtreffer in einer gemeinsamen Ergebnispräsentation aufbereitet werden [14, 16, 12, 23, 20, 21]. In ersten Realisierungsansätzen, wie bspw. bei den Suchdiensten von Google [1] oder Ask [2], wird dazu derart vorgegangen, dass die gewohnte und verbreitete Suchumgebung der Websuche als zentraler Interaktionspunkt zur Suche dient, worüber die gestellte Suchanfrage sowohl an die horizontale Websuchmaschine als auch an die verfügbaren vertikalen Suchmaschinen weitergeleitet wird. Wie im unteren Teil der Abbildung 1 verdeutlicht, werden die verschiedenen Typen von Informationsartefakten wie Bilder, Videos oder Newsbeiträge als Ergebnisse der unterschiedlichen vertikalen Suchdienste direkt mit in die Ergebnispräsentation der Websuche eingeblendet und hinsichtlich ihrer Relevanz geordnet [24, 16, 14]. Folglich bestehen die Einträge der Trefferpräsentation nunmehr nicht nur ausschließlich aus Links zu Webseiten oder Dokumenten, sondern es werden ebenfalls verschiedene Medientypen wie Videos, Landkarten oder Blogeinträge als explizite Suchtreffer mit einbezogen und visualisiert. Google geht dabei bspw. sogar so weit, dass Videos als Suchtreffer direkt in der Ergebnispräsentation betrachtet werden können [14].

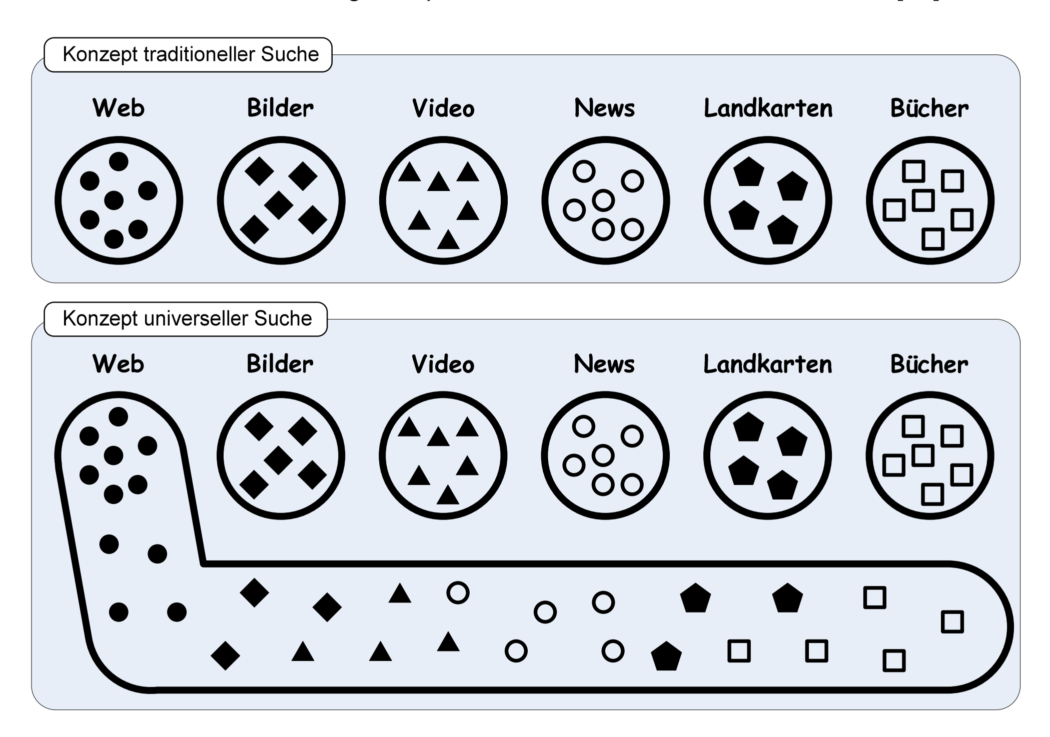


Abbildung 1: Gegenüberstellung von herkömmlicher und universeller Suche [16]

Durch den verfolgten integrativen Ansatz ermöglicht es das Konzept der universellen Suche, die jeweiligen Vorteile der horizontalen und vertikalen Suchdienste zu verbinden und darüber hinaus zusätzliche Synergieeffekte zu erzielen. Durch den Umstand, dass die verschiedenen vertikalen Suchdienste mit ihren spezifisch auf das fokussierte Themenfeld ausgerichteten Funktionalitäten und der damit verbundenen präziseren Kenntnis über die semantischen Bedeutungen von Information innerhalb der betrachteten Domäne direkt in die Websuche einbezogen werden, wird einerseits der Auswahlaufwand von entsprechend relevanten Suchdiensten und damit auch die Gefahr der unbewussten Vernachlässigung potenziell relevanter Suchdienste reduziert. Andererseits wird die Qualität des gesamten Suchdiensts im Hinblick auf die Präsentation von für die intendierte Suche möglichst relevanten Informationen erheblich verbessert [14, 12, 30]. Somit kann der Suchprozess vereinfacht und – damit verbunden – der Suchaufwand und die Suchkosten reduziert werden.

Dem Vorteil der Vereinfachung des Suchprozesses wirkt jedoch entgegen, dass die Trefferpräsentationen zunehmend mit multimedialen Informationsartefakten angereichert werden und somit das Risiko der „Verwaschung" von Suchergebnissen erwächst [12]. So werden bspw. zu der Suchanfrage *„Paris Hilton"* eine Vielzahl von Bildern, Videos oder Newsbeiträgen über die bekannte Hotelerbin ausgegeben, wobei dieselbe Suchanfrage auch zu einem Stadtplan der Stadt Paris mit darauf vermerkten Hotels führen kann. Darüber hinaus kann durch die Vielzahl verschiedener Medientypen die Übersichtlichkeit der Trefferpräsentation gemindert werden oder gar zu einem Interessenskonflikt führen, da evtl. gewisse Medientypen von der Suche auszuschließen sind, dies jedoch nicht direkt oder nur durch entsprechende Suchparameter in der Anfrage möglich ist. Aus diesem Grund kann eine wesentliche Problematik der universellen Suche in der angemessenen Aufbereitung der Trefferpräsentation hinsichtlich der Relevanz gesehen werden [13, 14].

Bei einer technischen Betrachtung des Konzepts der universellen Suche wird deutlich, dass dieses grundlegend auf dem Prinzip der Metasuche aufbaut [22, 23]. Metasuchmaschinen zeichnen sich im Wesentlichen durch das Merkmal aus, dass diese Suchanfragen parallel an eine Vielzahl anderer Suchmaschinen weiterleiten und die jeweiligen Suchergebnisse gesammelt in einer Ergebnispräsentation aufbereiten [6]. Das Konzept der universellen Suche arbeitet letztlich nach diesem Funktionsprinzip, indem von einer zentralen Suchoberfläche, vornehmlich der Websuche, eine Suchanfrage parallel an die verschiedenen Suchmaschinen weitergeleitet wird, die daraufhin ihren jeweiligen Suchindex auswerten und die jeweilig resultierenden Ergebnisse gesammelt in einer Trefferliste präsentiert werden.

Gegenüber dem Prinzip der Metasuche kann das Konzept der universellen Suche jedoch dahin gehend ergänzt werden, dass sich die verschiedenen vertikalen Suchmaschinen, die im Rahmen der universellen Suche angefragt werden, aufgrund der Ausrichtung auf die speziellen Gegebenheiten der fokussierten Domäne, in ihrer grundlegenden Funktionalität unterscheiden. Schon der Aufbau eines Suchindex für die unterschiedlichen Medientypen wie Videos, Blogs oder Landkarten bedarf unterschiedlicher algorithmischer und technischer Konzepte, wie bspw. Crawler, um auf die unterschiedlichen Medientypen zugreifen und diese in den Suchindex aufnehmen zu können [13]. Darüber hinaus bedarf es ebenfalls verschiedener Konzepte der Trefferpräsentation, die an den jeweiligen Besonderheiten der unterschiedlichen Medientypen angepasst sind [24]. In Anbetracht der speziellen funktionalen und technischen Eigenschaften der verschiedenen vertikalen Suchmaschinen wird die Notwendigkeit einer gesonderten Integration dieser in die Umgebung der universellen Suche deutlich. Aus diesem Grund beschränken sich bestehende Anbieter der universellen Suche wie Google [1] oder Ask [2] derzeit mehrheitlich auf die Integration von ausgewählten eigenen vertikalen Suchdiensten.

Mit der Integration der Suchergebnisse aus den verschiedenen vertikalen Suchmaschinen in eine gemeinsame Ergebnispräsentation erwächst die Problematik der Bestimmung der Relevanz von unterschiedlichen Medientypen [13, 20]. Für die Bestimmung der Relevanz von Webseiten nutzt bspw. Google mit dem PageRank einen Algorithmus, der anhand von eingehenden Links und dem Gewichtungsfaktor der verweisenden Webseite rekursiv das Gewicht einer Webseite als Maßzahl für die Relevanz errechnet [31]. Die Anwendung dieses Algorithmus ist aufgrund der speziellen Eigenschaften verschiedener Medientypen bspw. bei der Bilder- oder Videosuche nicht möglich. Somit wird deutlich, dass den verschiedenen vertikalen Suchmaschinen unterschiedliche Bewertungssysteme und -skalen für die Relevanz der Suchtreffer zu-

grunde liegen. Da ein vordefiniertes Schema, welches eine Rangfolge der Relevanz verschiedener Medientypen vorgibt, in diesem Zusammenhang nicht realisierbar ist, bedarf es einer algorithmischen Methode, die eine Normalisierung der unterschiedlichen Bewertungsschemata und -skalen ermöglicht [13, 20]. Dies ist die grundlegende Voraussetzung um die unterschiedlichen Suchergebnisse hinsichtlich ihrer Relevanz gegenüberstellen und entsprechend sinnvoll in einer einheitlichen Ergebnispräsentation zusammenführen zu können. Im Zusammenhang mit den verschiedenen Realisierungsbestrebungen der universellen Suche hat sich die Normalisierung als komplexe Problemstellung erwiesen, wobei es nach bisherigen Erkenntnissen nur dem Suchmaschinenbetreiber Google gelungen ist, eine entsprechende Methode dazu zu entwickeln [13]. Mit dieser Entwicklung wurde bei Google ein grundlegendes Problem der universellen Suche gelöst und damit ein wesentlicher Fortschritt im Hinblick auf die Realisierung des Konzepts der universellen Suche errungen.

3 Betrachtung aus ökonomischer Perspektive

Als einer der Bereiche, der die Interneteuphorie und Dotcom-Blase überstanden hat, hat sich der Betrieb von Suchmaschinen in den letzten Jahren zu einem eigenständigen Markt und bedeutenden ökonomischen Faktor entwickelt. So wird das Volumen des globalen Marktes von Suchdiensten im Jahre 2006 auf ca. 14 Mrd. $ geschätzt und für das Jahr 2007 ein Wachstum von 31 % prognostiziert [11]. Auf den mit derzeit fast 60 % Marktanteil führenden Suchmaschinenbetreiber Google entfällt davon ein Umsatzvolumen im Jahr 2006 von rund 3,2 Mrd. $ [10].

Den mehrheitlichen Anteil des Umsatzes generieren Suchmaschinenbetreiber derzeit durch Advertisingkonzepte [10]. Diese ermöglichen es, Werbebotschaften oder Anzeigen bspw. auf Basis von verkauften Suchwörtern zusätzlich zu den jeweiligen Suchergebnissen einzublenden [32]. Suchmaschinen erweisen sich somit als probates Konzept des Direktmarketings, indem sie durch die themenbezogene Platzierung der Werbung eine zielgerichtete Ansprache von Kunden ermöglichen. Unternehmen wird es damit ermöglicht, die Ausgaben für Werbemaßnahmen gezielter platzieren und damit eine entsprechend höhere Wirkung erzielen zu können. Durch die gezielte Präsentation von Angeboten wird nicht zuletzt aufseiten des Nutzers von Suchdiensten der Suchprozess vereinfacht und damit der Aufwand und die Kosten der Suche verringert. Idealistisch betrachtet kann somit die Attraktivität des Suchdienstes gesteigert werden, was nicht zuletzt für eine Erweiterung des Nutzerkreises förderlich ist und damit ebenfalls zur Steigerung der Attraktivität des Suchdienstes als mögliche Werbeplattform beiträgt.

Bei einer Übertragung dieses Gedankens auf das Konzept der universellen Suche werden weitere Potenziale sowohl für Suchmaschinenbetreiber als auch für andere Akteure deutlich. Durch den direkten Einbezug von spezialisierten Suchmaschinen in die Suche erschließen sich neue Medientypen, worüber spezielle Werbebotschaften u. U. zielgerichteter positioniert werden können. So bietet sich bspw. die Möglichkeit, Anzeigen über geografische Suchdienste wie bspw. Google Maps [9] zu positionieren und dadurch die angesprochene Zielgruppe über regionale Kriterien zielgerichtet anzusprechen. Dieses Konzept bietet sich vor allem für Unternehmen im Hotel- und Gaststättengewerbe an [33]. Eine weitere Möglichkeit, spezialisierte Suchdienste für die zielgerichtete Ansprache von potenziellen Kunden zu nutzen, sind Produktsuchmaschinen. Derartige Suchdienste erlauben es, angebundene Kataloge oder Onlineshops direkt zu durchsuchen und daraus entsprechende Produktangebote zu präsentieren. Auf diesem Wege können

einerseits Onlineshops, die u. U. einen kleinen Nischenbereich abdecken, einen breiteren Absatzmarkt erschließen bzw. einen größeren Kundenkreis ansprechen und andererseits wird eine zielgerichtete Suche nach Produkten aus Sicht des Nutzers ermöglicht [33].

In Anbetracht der mit dem Konzept der universellen Suche verbundenen Potenziale sind vor allem für Suchmaschinenbetreiber neue Geschäftsmodelle denkbar. Durch die Verbreiterung der Suchbasis um neue Medientypen haben sich zusätzliche Möglichkeiten für die Positionierung von Werbebotschaften und Anzeigen ergeben, was nicht zuletzt neue Einnahmequellen für Suchmaschinenbetreiber eröffnet. Darüber hinaus ist es denkbar, dass Suchmaschinenbetreiber zusätzlich entgeltpflichtige Suchdienste zu Verfügung stellen, die auf sehr spezielle Domänen, wie bspw. Onlineshops, Communityportale, Wissensdatenbanken oder einzelne Webauftritte von Unternehmen, fokussiert sind. Die Entrichtung des Nutzungsentgelts kann dabei auf den Betreiber der Suchbasis oder auf den Nutzer des speziellen Suchdienstes übertragen werden. Wird dies dem Nutzer übertragen, könnten diesem verschiedene Varianten universeller Suche angeboten werden, die unterschiedliche spezialisierte Suchdienste umfassen.

Bei einer grundlegenden ökonomischen Betrachtung dieser Geschäftsmodelle kommt jedoch die Fragestellung auf, warum Unternehmen Werbeangebote über Medientypen schalten sollten, die nur über spezialisierte Suchdienste erreichbar sind bzw. warum sich u. U. der Betrieb stark spezialisierter Suchdienste auszahlen könnte [25]. Diese Fragestellung lässt sich anhand des von Chris Anderson als *Long-Tail-Effekt* [7, 8] bezeichneten Phänomens beantworten. Im Zusammenhang mit der Suche im Web spiegelt sich dieser Effekt derart wider, dass nur eine geringe Auswahl an Suchanfragen mit einer hohen Anfragehäufigkeit existiert, die Anfragehäufigkeit der mehrheitlichen Suchanfragen jedoch bis auf ein gewisses Niveau stark abfällt [27, 25].

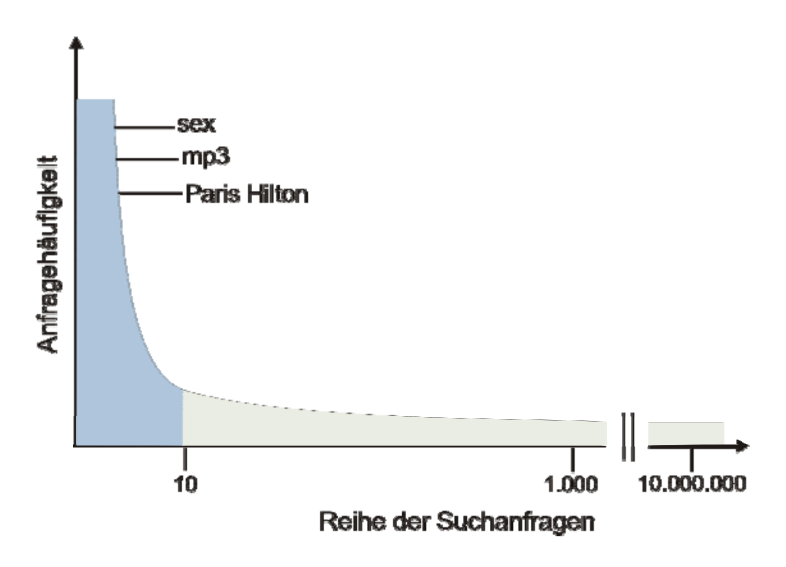


Abbildung 2: Long-Tail-Effekt von Suchanfragen [25]

Wie in Abbildung 2 dargestellt, entsteht somit eine auf einem Niveau lang auslaufende Kurve, die einen gewissen Schwellenwert nicht unterschreitet. Gemäß dem Verständnis des Long-Tail-Effekts spiegelt sich der bedeutendste ökonomische Nutzen für die universelle Suche gerade in den Suchanfragen wider, die sich in diesem auslaufenden Nischenbereich befinden [25]. Durch die Verbindung einer Vielzahl an spezialisierten Suchdiensten mit der breiten Websuche ermöglichen entsprechende Suchmaschinen einen Zugriff auf ein wesentlich breiteres Spektrum an Informationsquellen und -artefakten und decken damit weitestgehend den gesamten Nischen-

bereich des „Long-Tail" ab. Somit wird deutlich, dass im Zusammenhang mit dem Konzept der universellen Suche sich der Betrieb von spezialisierten Suchdiensten sowohl für Suchmaschinenbetreiber als auch für werbende und suchende Nutzer aus ökonomischer Perspektive ebenfalls als vorteilhaft erweisen kann.

4 Betrachtung aus sozialer Perspektive

Suchen im Web wird in heutigem Sprachgebrauch oftmals als *„googeln"* bezeichnet. Dies ist jedoch nicht zuletzt auf den hohen Bekanntheitsgrad des Suchdienstes zurückzuführen. Im Umkehrschluss deutet der hohe Bekanntheitsgrad letztlich auf eine intensive und verbreitete Nutzung eines derartigen Dienstes hin. Ebenso wie die Bezeichnung eines Suchdienstes für die einfache Tätigkeit des Suchens immer allgegenwärtiger im täglichen Sprachgebrauch wird, nimmt auch die Tätigkeit des Suchens im Web eine immer allgegenwärtigere Stellung ein. Sie wird mehr und mehr zu einer Selbstverständlichkeit bzw. erfolgt zunehmend unbewusst. Bieten Suchdienste wie die universelle Suche einen einfachen und breiten Zugriff auf verschiedene Informationsangebote und Medientypen an, kann dies sogar dazu führen, dass in einigen Situationen des täglichen Lebens traditionelle Gewohnheiten geändert oder gar ersetzt werden. So ist es denkbar, dass ein Blick in die Onlineenzyklopädie Wikipedia [5] den Griff ins Bücherregal zu einem Lexikon ersetzen könnte. Wird dieser Gedanke in überspitztem Maße weiter vorangetrieben, hätte dies u. U. grundlegende kulturelle Veränderungen zur Folge [25].

Ein erster Schritt zur Änderung der Kultur hinsichtlich der Nutzung des Webs wurde zunehmend durch den Gedanken des *„Mitmach-Webs"* proklamiert [27]. Die Nutzer von Informationsangeboten im Web nehmen nunmehr nicht nur eine konsumierende, sondern vielmehr eine produzierende Rolle ein. Benutzer können eigene Informationsangebote wie Blogs aufbauen und Informationen für andere Benutzer bereitstellen. Sie können jedoch auch bestehende Informationsartefakte wie Videos oder Bilder durch Kommentare beschreiben und diese somit implizit in einen Kontext einordnen. Durch diese Beschreibung werden die Informationsartefakte mit zusätzlichen Metadaten versehen, die nicht nur von anderen Benutzern ausgewertet werden können, sondern auch von Suchdiensten. Dies kommt vor allem bei multimedialen Informationsartefakten zum Tragen, die von Suchmaschinen fast ausschließlich anhand der Metadatenbeschreibung als relevant für einen Suchkontext identifiziert werden können. Vor allem für die auf multimediale Informationsartefakte ausgerichteten Suchdienste und damit verbunden auch für das Konzept der universellen Suche eröffnet sich durch die Änderung der Nutzungskultur im Web hin zur Partizipation eine Reihe von Vorteilen und Möglichkeiten. Diese führen nicht zuletzt zur Steigerung der Qualität der Suchdienste in Bezug auf die Präsentation relevanter Informationen.

Die zunehmende partizipative Rolle von Benutzern im Web birgt im Zusammenhang mit der Verbreitung universeller Suchdienste jedoch gewisse Gefahren. Durch eine Vielzahl von Angeboten wie Blogs, Video- oder Bilderportale werden Nutzer geradezu angeregt, Informationen zu publizieren. Dabei werden nicht nur allgemeine Informationen bereitgestellt, sondern häufig auch unbewusst persönliche Informationen. Die Möglichkeit einer derartigen Erweiterung des Suchspektrums, wie es durch universelle Suchdienste angestrebt wird, verschafft somit zunehmend einen breiten Zugriff auf persönliche Daten und birgt die Gefahr des Verlusts der Privatsphäre.

5 Ausgewählte Realisierungen

Im Umfeld der marktanteilsmäßig größten Suchmaschinenbetreiber bestehen verschiedentlich Bestrebungen das Konzept der universellen Suche umzusetzen. Dies folgt nicht zuletzt dem Gedanken des Alleinstellungsmerkmals, welches sich Suchmaschinenbetreiber dadurch verschaffen würden. Neben Google [1] versuchen auch bedeutende Suchmaschinenbetreiber wie Yahoo! [2], Ask [3], oder Microsoft [4] eine entsprechende universelle Suchumgebung zu realisieren [17, 19]. Wie der universelle Suchdienst von Google und dem gegenüber die Suchumgebung Ask X realisiert wurde, ist Gegenstand einer kurzen Betrachtung.

Google Universal Search

Auf der Anfang dieses Jahres stattgefundenen Konferenz „*Searchnomics*“ hat Marissa Mayer, Vizepräsidentin des Bereichs Search Products bei Google, die realisierte *Google Universal Search* vorgestellt [15, 14, 16, 21]. Mit der weit verbreiteten Websuche und der Vielzahl eigener vertikaler Suchdienste bietet sich Google eine entsprechend gute Ausgangbasis für die Realisierung der universellen Suche. Dem grundlegenden Konzept der universellen Suche (vgl. Kapitel 2) folgend, nutzt Google die eigene Websuche und bindet derzeit jedoch nur die verschiedenen eigenen spezialisierten Suchdienste an die Suchoberfläche an, sodass sämtliche Suchdienste über das zentrale Suchfeld der Websuche simultan angefragt werden können. Die Metapher der einzelnen Tabs für die speziellen Suchdienste wie Bilder- oder Videosuche wird somit implizit auf das zentrale Suchfeld übertragen und für den Benutzer transparent gestaltet. Um eine gezielte Suche dennoch zu ermöglichen, bietet Google weiterhin entsprechende Tabs für den direkten Zugriff auf die spezialisierten Suchdienste an.



Abbildung 3: Beispiel der Google Universal Search

Die Integration und Präsentation der Suchtreffer aus den verschiedenen Suchdiensten erfolgt bei Google in Form der traditionellen Trefferauflistung der Websuche. In Abhängigkeit der Relevanz werden somit die Treffer der unterschiedlichen Medientypen wie Bilder, Newsbeiträge oder Landkarten in der Trefferauflistung direkt zwischen den Links auf Webseiten oder Doku-

menten eingeblendet (vgl. Abbildung 3). Bei multimedialen Treffern wie Bilder Videos oder Landkarten werden diese innerhalb der Trefferliste direkt visualisiert. Bei der Präsentation von Videos ermöglicht es Google darüber hinaus, dass – wie in Abbildung 3 dargestellt – Videos direkt in der Trefferauflistung betrachtet werden können. Treten in der Trefferauflistung mehrfach Einträge eines Medientyps mit nahezu identischer Relevanz auf, so werden diese zur Wahrung der Übersicht gruppiert dargestellt. Darüber hinaus bietet Google die Möglichkeit eines „drill-downs“ der Trefferauflistung an, sodass diese hinsichtlich einzelner Medientypen gezielt gefiltert werden kann. Somit ermöglicht es Google, die integrierte Ergebnispräsentation der universellen Suche wiederum auf die Ergebnismengen einzelner spezialisierter Suchen herunterzubrechen.

Ask X

Ebenfalls einen Suchdienst für die universelle Suche im Web hat der Suchmaschinenbetreiber Ask umgesetzt und betreibt diesen derzeit prototypisch unter der Bezeichnung Ask X [2]. Vergleichbar mit Google nutzt Ask X derzeit ebenfalls die ursprüngliche Suchoberfläche der Websuche und bindet für den Nutzer transparent im Hintergrund bisher nur die eigenen vertikalen Suchdienste an [18, 17]. Gegenüber dem Marktführer Google bietet Ask X jedoch nur einen geringeren Umfang an vertikalen Suchdiensten an, die sich in ihrer Art nicht von den spezialisierten Suchdiensten von Google abheben.

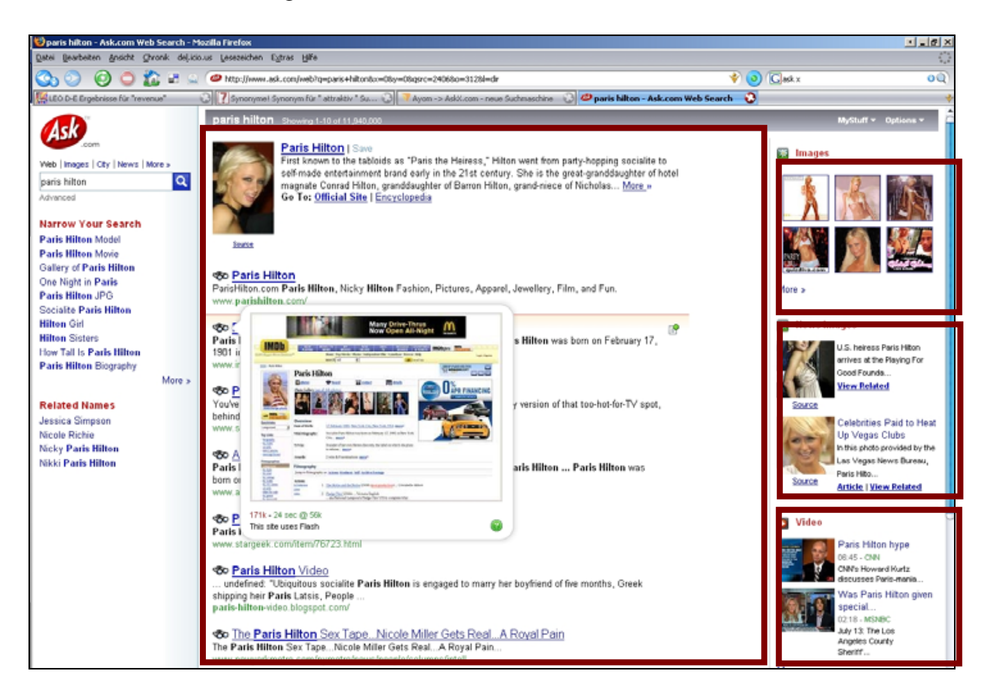


Abbildung 4: Beispiel der universellen Suche Ask X

Gegenüber der Google Universal Search unterscheidet sich Ask X jedoch im Konzept der Integration und Präsentation der Suchergebnisse aus den unterschiedlichen Suchdiensten. Entgegen der von Google verfolgten Auflistung der gesamten Suchergebnisse in einer Trefferliste, erfolgt die Präsentation der Suchergebnisse bei Ask X – wie in Abbildung 4 verdeutlicht – in verschiedenen Spalten je nach Medientyp gruppiert. Eine zentrale Stellung nehmen dabei die Suchergebnisse aus der traditionellen Websuche ein, die – wie in Abbildung 4 dargestellt – als separate und größenmäßig dominierende Spalte mittig platziert werden. Die Ergebnisse der verschiedenen vertikalen Suchdienste werden hingegen in einer kleineren Spalte am rechten

Rand je nach Medientyp gruppiert präsentiert. Die starke Trennung der speziellen Medientypen fördert einerseits zwar die Übersichtlichkeit der Trefferpräsentation, entspricht jedoch nicht in vollem Maße dem Gedanken der universellen Suche. Durch die strikte Trennung obliegt es somit dem Benutzer, die Relevanz zwischen den einzelnen Suchergebnissen der verschiedenen Medientypen zu bestimmen. Dies kann letztlich darauf zurückgeführt werden, dass in Ask X derzeit kein Verfahren zur Normalisierung der unterschiedlichen Bewertungsskalen implementiert ist, um eine Vergleichbarkeit zwischen den Suchergebnissen der unterschiedlichen Medientypen zu ermöglichen. Analog zur universellen Suche von Google ermöglicht es Ask X ebenfalls, die Trefferliste auf die Präsentation einzelner Medientypen herunterzubrechen.

6 Universelle Suche – Die perfekte Suche?

Als aktueller Entwicklungstrend von Suchdiensten im Web wird derzeit das Konzept der universellen Suche diskutiert und hält schrittweise Einzug in bestehende Suchdienste wie Google oder Ask. Dem Grundgedanken der Metasuche folgend, versuchen Suchmaschinenbetreiber mit dem Konzept der universellen Suche die breit diversifizierte Landschaft spezialisierter Suchdienste mit dem traditionellen Dienst der Websuche unter einem gemeinsamen Suchdienst zu verknüpfen und somit aus der Verbindung der jeweiligen Vorteile der beiden Konzepte weitere Synergieeffekte zu erzielen. Als positive Effekte werden dabei neben der Möglichkeit eines einfacheren Zugriffs auf ein breiteres Spektrum an Informationen vor allem die Verbesserung der Relevanz präsentierter Suchtreffer genannt, was im Allgemeinen auch als wesentliches Qualitätsmerkmal von Suchdiensten angesehen wird.

Eine Betrachtung dieser Effekte lässt schnell die Vermutung aufkommen, dass das Konzept der universellen Suche die Problemstellung der bisherigen Ansätze beheben kann und damit den Bereich der Websuche revolutioniert. Bei einer genaueren Betrachtung dieses Konzeptes wird jedoch deutlich, dass universelle Suchdienste allein nur einen Beitrag leisten können, um den grundsätzlichen Problemstellungen der Suche im Web zu begegnen. Schon dem grundlegenden Problem der Suche im Web – die Identifikation der perfekten Treffer – kann von universellen Suchdiensten nur bedingt begegnet werden, indem durch ein verbreitertes Suchspektrum zwar der Zugriff auf die mögliche Menge potenziell relevanter Informationen verbessert wird, jedoch der Anteil relevanter Informationen in der Treffermenge nur bedingt erhöht bzw. u. U. sogar verringert werden kann [26]. Um dieser Herausforderung zu begegnen, werden aktuell ebenfalls Ansätze zur Personalisierung von Suchdiensten [34], zur zeitraumbezogenen Analyse von Suchanfragen oder den Einbezug von Semantic-Webkonzepten diskutiert [25].

Im Hinblick auf die Entwicklung der perfekten Suche für das Web stellt sich jedoch die Frage, wie diese verschiedenen Suchkonzepte in Verbindung genutzt werden können, um eine perfekt personalisierte, kontextbezogene und zielgerichtete Suche zu ermöglichen. Die Antwort auf diese Fragestellung ist überraschend einfach, eine Kombination der Ansätze wird eine perfekte Suche nicht ermöglichen, sondern erlaubt nur einen weiteren Schritt in diese Richtung. Die Realisierung des perfekten Suchdienstes wird somit wohl immer eine Wunschvorstellung bleiben. Ob es wirklich an dem ist, wäre jedoch interessant zu erfahren [25].

Referenzen

[1] Google, http://www.google.com. Zuletzt abgerufen am 06.09.2007.
[2] Ask X, http://www.askx.com. Zuletzt abgerufen am 06.09.2007.
[3] Yahoo! Alpha, http://au.alpha.yahoo.com. Zuletzt abgerufen am 05.09.2007.
[4] Microsoft Live Search, http://search.imagine-live.com. Zuletzt abgerufen am 05.09.2007.
[5] Wikipedia, http://de.wikipedia.org/wiki/Hauptseite. Zuletzt abgerufen am 01.09.2007.
[6] Wikipedia: Metasuchmaschine, http://de.wikipedia.org/wiki/Metasuche. Zuletzt abgerufen am 01.09.2007.
[7] Wikipedia: Long Tail, http://de.wikipedia.org/wiki/Long_Tail. Abgerufen am 04.09.2007.
[8] Wired 12.10: The Long Tail, http://www.wired.com/wired/archive/12.10/tail.html. Abgerufen am 04.09.2007.
[9] Google Maps, http://maps.google.com. Zuletzt abgerufen am 02.09.2007.
[10] Google Investor Relations, http://investor.google.com/releases/2006Q4.html. Zuletzt abgerufen am 05.09.2007.
[11] SearchEngineWatch.com: Global Search Market, http://blog.search-enginewatch.com/blog/060104-161552/. Zuletzt abgerufen am 05.09.2007.
[12] Nexi Information Technology Blog: Suchmaschinenoptimierung für Google Universal Search, http://de.nexi.is.it/blog/index.php/2007/06/29/suchmaschinenoptimierung-fuer-google-universal-search/. Zuletzt abgerufen am 02.09.2007.
[13] SearchEngineWatch.com: Will universal Search Mean Universal Domination?, http://searchenginewatch.com/showPage.html?page=3625897. Zuletzt abgerufen am 05.09.2007.
[14] SearchEngineLand.com: Google 2.0: Google Universal Search, http://searchengineland.com/070516-143312.php. Zuletzt abgerufen am 29.08.2007.
[15] ReadWriteWeb.com: Google´s Marissa Meyer on the Future of Search, http://www.readwriteweb.com/archives/google_marissa_mayer_future_of_search.php. Zuletzt abgerufen am 28.08.2007.
[16] SearchEngineLand.com: Search Illustrated: Google´s Universal Search Explained, http://searchengineland.com/070703-084856.php. Zuletzt abgerufen am 02.09.2007.
[17] ReadWriteWeb.com: How Does Google Universal Search Compare to Yahoo! Alpha, Ask X, Microsoft Image Life & Google´s Own Searchmash?, http://www.readwriteweb.com/archives/how_does_google_-universal_search_compare.php. Zuletzt abgerufen am 30.08.2007.
[18] ReadWriteWeb.com: Ask3D: Ask.com One-Ups Google universal Search, http://www.readwriteweb.com/archives/ask3d.php. Zuletzt abgerufen am 06.09.2007.
[19] ReadWriteWeb.com: how Alt Search Engines implemented Universal Search…Well Before Google Did, http://www.readwriteweb.com/archives/-how_alt_search_engines_implemented_universal_search.php. Zuletzt abgerufen am 30.08.2007.
[20] googleblog.blogspot.com: Behind the scenes with universal search, http://googleblog.blogspot.com/2007/05/behind-scenes-with-universal-search.html.Zuletzt abgerufen am 25.08.2007.
[21] googlesystem.blogspot.com: Google Launches New Design, Universal Search an a Site for Search Experiments, http://googlesystem.blogspot.com/search/label/Universal %20Search. Zuletzt abgerufen am 05.09.2007.

[22] resourceshelf.com: Let´s Talk Metasearch, Federated Search, or Universal Search, http://www.resourceshelf.com/2007/05/16/lets-talk-federated-or-metasearch/. Zuletzt abgerufen am 20.08.2007.

[23] resourceshelf.com: Search engine mashup or What Some Call Federated or Meta Searching, http://www.resourceshelf.com/2007/07/06/search-engine-mashup-or-what-some-call-federated-or-meta-searching/. Zuletzt abgerufen am 05.09.2007.

[24] Lewandowski, Dirk: Trefferpräsentation in Web-Suchmaschinen, in: Wissenschaft, Bildung und Wirtschaft; proceedings of the 29th Online-Tagung der DGI 2007.

[25] Battelle, John: The Search – How Google and Ist Rivals Rewrote the Rule of Business and Transformed Our Culture. Portfolio, New York 2005.

[26] Baeza-Yates, Ricardo; Ribeiro-Neto, Berthier: Modern Information Retrieval. Addison-Wesley, New York 1999.

[27] Vossen, Gottfried; Hagemann, Stephan: Unleashing Web 2.0 – From Concepts to Creativity. Elsevier, Burlington 2007.

[28] Wikipedia: Suchmaschine, http://de.wikipedia.org/wiki/Suchmaschine. Zuletzt abgerufen am 05.09.2007.

[29] Wikipedia: Vertical Search, http://en.wikipedia.org/wiki/Vertical_search/. Zuletzt abgerufen am 05.09.2007.

[30] ReadWriteWeb.com: Google Universal Search – is Vertical Search Space Finished?, http://www.readwriteweb.com/archives/google_universal_-search_vertical_search_finished.php. Zuletzt abgerufen am 05.09.2007.

[31] Wikipedia: PageRank, http://en.wikipedia.org/wiki/Page_rank. Zuletzt abgerufen am 05.09.2007.

[32] Wikipedia: Keyword Advertising, http://en.wikipedia.org/wiki/Keyword_advertising. Zuletzt abgerufen am 05.09.2007.

[33] SearchEngine Roundtable: How to Capitalize on Google Universal Search, http://www.seroundtable.com/archives/013594.html. Zuletzt abgerufen am 05.09.2007.

[34] webpronews.com: Is Personalized Search the Future?, http://www.webpronews.com/topnews/2004/08/03/is-personalized-search-the-future. Zuletzt abgerufen am 06.09.2007.

[35] Wikipedia: Deep Web, http://de.wikipedia.org/wiki/Deep_Web. Zuletzt abgerufen am 10.09.2007.

Spezialisiertes Suchen im Web

Felix Müller-Wienbergen

1 Von der generischen zur spezialisierten Suche

Das World Wide Web ist mittlerweile unumstritten ein prägendes Element unserer Gesellschaft geworden und Suchmaschinen besitzen die Aufgabe, alltäglich den bequemen Zugang zu der schier unüberschaubaren Fülle an Informationen im Web zu gewähren. Als ein Indiz auf die zentrale Rolle, die Suchmaschinen heutzutage einnehmen, kann die Namensgebung des Webdienstes Google Zeitgeist[1] angeführt werden. Durch das reine Auszählen der beliebtesten Suchterme in bestimmten Perioden nimmt der Dienst für sich in Anspruch, eine Chronologie von Spiegelbildern der Gesellschaft/des Zeitgeistes zu präsentieren [Bat05].

Bei diesen alltäglichen Suchprozessen ist die Intention der Nutzer großer Suchmaschinen wie Google, Yahoo! und MSN nicht immer identisch. ROSE und LEVINSON identifizieren im Groben drei unterschiedliche Intentionen für die Verwendung einer Suchmaschine [RL04]:

- *Navigational:* Nutzer einer Suchmaschine, die diese Intention verfolgen, suchen explizit die Homepage einer bestimmten Institution oder Firma. Sie haben eine ganz konkrete Webseite im Kopf und nutzen die Suchmaschine nur, weil sie entweder die URL nicht wissen oder es bequemer ist, nach dieser zu suchen als sie vollständig einzugeben.
- *Informational:* Suchanfragen mit diesem Fokus zielen darauf ab, Informationen zu einem bestimmten Thema zu erhalten. Nutzer mit einer informationalen Intention müssen entweder eine bestimmte Frage beantworten, benötigen Hilfestellung bei einem bestimmten Sachverhalt oder wollen generell etwas zu einem bestimmten Themengebiet lernen.
- *Ressource:* Bei der Gruppe von Intentionen, die ROSE und LEVINSON unter dem Begriff Ressource zusammenfassen, geht es darum, etwas Anderes als reine Information zu erhalten. Hierbei kann es sich um den Download digitaler Inhalte wie Computerprogramme oder Musikdateien handeln, die einen Computer benötigen, um verwendet werden zu können. Der Besuch von Webseiten zu reinen Unterhaltungszwecken fällt ebenfalls in diese Kategorie. Zudem zählen interaktive Dienste, die Eingaben des Nutzers erfordern, wie z. B. digitale Kartendienste, Wettervorhersagen oder Onlineshops, zur Ressourcenintention. Ebenfalls hinzuzuzählen ist die Akquise von Artefakten, die nicht nur am Computer verwendet werden können, sondern ebenfalls in der „Offlinewelt" Nutzen stiften. Hierunter fällt der Download von bspw. Liedtexten, Rezepten oder anderen Dokumenten, die ausgedruckt als materielle Ressourcen gehandhabt werden können.[2]

---

1 http://www.google.ch/intl/de/press/zeitgeist.html. Mittlerweile wird die Historie auf dieser Seite nicht weiter fortgeschrieben. Stattdessen hat Google den Service in seinen neuen Dienst *Hot Trends* integriert (http://www.google.com/trends/hottrends).

2 BRODER stellte einen ähnlichen, weniger detaillierten Ordnungsrahmen auf [Bro02]. Abweichend von ROSE und LEVINSON nennt er die dritte Klasse von Suchintentionen *transaktional* und nicht Ressource.

Um zu bewerten, inwieweit eine Suchmaschine die jeweilige Intention eines Nutzers befriedigt, kann in Anlehnung an die Evaluation von Information Retrieval Systemen die Effektivität anhand zweier Kennzahlen bestimmt werden. RIJSBERGEN schlägt zu diesem Zweck die beiden Messgrößen *Abruf* und *Präzision*[3] vor. Mit ihrer Hilfe wird die Fähigkeit eines Systems bemessen, relevante Informationen zu präsentieren und gleichzeitig irrelevante Informationen aus der Ergebnismenge herauszufiltern. Es wird davon ausgegangen, dass, je effektiver ein System in dieserlei Hinsicht ist, desto besser befriedigt es die Intentionen seiner Nutzer [Rij79].

	Relevant	Nicht-Relevant	
2-3 Präsentiert	$A \cap B$	$\bar{A} \cap B$	B
2-3 Nicht-Präsentiert	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	$\bar{B}$
2-3	A	$\bar{A}$	N

Sei N die Menge aller möglichen Informationen, die eine Suchmaschine auf eine Suchanfrage hin präsentieren könnte. Diese Menge teilt sich auf in die vier paarweise disjunkten Teilmengen der relevanten A- und nicht-relevanten $\bar{A}$- bzw. der präsentierten B- und nicht-päsentierten $\bar{B}$-Informationen. Bspw. wäre für eine navigationale Suchanfrage $A=1$ *und* $\bar{A}=N-1$. Auf dieser Basis ist Präzision definiert als das Verhältnis von relevanten Informationen, die ein System auf eine Suchanfrage hin zurückgeliefert hat, zu der Gesamtmenge an Informationen, die präsentiert werden (vgl. Ausdruck 1). Demgegenüber ist der Abruf definiert als das Verhältnis relevanter präsentierter Informationen zu existenten relevanten Informationen (vgl. Ausdruck 2) [Rij79].

$$\text{Präzision} = \frac{|A \cap B|}{|B|} \quad (1)$$

$$\text{Abruf} = \frac{|A \cap B|}{|A|} \quad (2)$$

Viele Webuser nutzen als Ausgangspunkt einer Suche eine der größten drei Suchmaschinen – Google, Yahoo! und MSN – und starten dort eine generelle Suche. Diese Suchdienste gehören zu der Klasse der generischen Suchmaschinen. Generische Suchmaschinen decken ein möglichst breites Spektrum an thematischen Suchgebieten ab, indem ihr Index mit allem befüllt wird, was in die Reichweite zugehöriger Crawlerprozesses gelangt [Lar07]. Sie werden daher auch als horizontale Suchmaschinen bezeichnet [Sul07]. Eigentlich versucht jedoch jeder Nutzer einer Suchmaschine, ein ganz spezielles Problem zu lösen, wie z. B. den Kauf eines Autos, die Suche nach einem günstigen Angebot für ein spezielles Computerzubehör, die Buchung einer Reise oder die Suche nach detaillierten Informationen zu einem aktuellen Geschehnis. Eben dieser Kontext, in dem eine Suche stattfindet, wird von spezialisierten Suchmaschinen berücksichtigt.

Spezialisiertes Suchen stellt eine verhältnismäßig neue Disziplin im Bereich der Websuche dar [Sok06]. Die Kernidee, die hinter dieser Art von Suchmaschinen steckt, ist die Reduzierung des durchsuchten Teils des Webs auf eine kleine thematisch relevante Teilmenge, eine a priori gegebene Nische [Mit07a]. Die Idee der spezialisierten Suche ist dabei nicht neu, jedoch nimmt

---

[3] Im Originaltext wird von *recall* und *precision* gesprochen.

erst in jüngster Zeit die Verbreitung dieser Suchdienste dramatisch an Geschwindigkeit zu [Kop05]. So werden gegenwärtig viele Dienste und Technologien zur gezielten Suche nach bestimmten Content-Formen und domänenspezifischen Informationen als „die Top-Innovationen" im Bereich der Internetsuche gehandelt [Kar07].

Verglichen mit generischen Suchmaschinen brillieren spezialisierte Suchdienste in ihrer jeweiligen thematischen Nische. Durch die thematische Vorselektion potenzieller Suchergebnisse N wird der Anteil angezeigter, nicht-relevanter Informationen $\bar{A} \cap B$ geringer. Wird bspw. an einen Suchdienst spezialisiert auf Personen eine Anfrage nach „Webentwicklung" gestellt, so werden unter den gelieferten Suchergebnissen keine Referenzen auf Fachliteratur zum Erlernen von Webentwicklung geliefert [Mit07a]. Den gleichen positiven Effekt auf die Ausblendung irrelevanter Informationen hat die Offerierung domänenspezifischer Suchparameter. Durch das A-priori-Wissen über die Charakteristika der potenziellen Ergebnismenge können Suchanfragen mithilfe entsprechender Attribute sehr präzise formuliert werden.[4] So können bspw. auf einem Gebrauchtwagenportal gesuchte Kraftfahrzeuge sehr detailliert über die Angabe von Marke, Farbe, Leistung, Baujahr und Kilometerlaufleistung spezifiziert werden. Zudem nehmen spezialisierte Suchmaschinen generell eine tiefere Erfassung des Webs in Bezug auf ihren semantischen Fokus vor, da sie Zugriff auf Ressourcen besitzen, die generische Suchmaschinen häufig außer Acht lassen. Diese indizieren vornehmlich statische Webseiten und sind nicht an dynamische Datenbanken mit bestimmten thematischen Schwerpunkten angeschlossen [Kop05, Kuc05]. Die Menge der relevanten und auch präsentierten Informationen $A \cap B$ ist folglich bei vertikalen Suchdiensten absolut größer [Ste01]. Spezialisierte Suchmaschinen sind daher in ihrer spezifischen Nische effektiver als generische Allzwecksuchmaschinen – sowohl im Hinblick auf ihre Präzision als auch des Abrufs. Sie eignet sich daher explizit für Suchen mit informationaler Intention bzw. Intentionen aus der Klasse Ressource. Für navigationale Suchanfragen fehlt ihnen jedoch per Definition die nötige Breite.

Als weiteren Vorteil spezifischer Suchmaschinen eröffnet das A-priori-Wissen über den Inhalt der Suchergebnisse die Möglichkeit, diese für den Nutzer besser visuell aufzubereiten. So ist die entsprechend der Schachtelung eingerückte Darstellung von Quellcode sicherlich ansprechender und besser zu erfassen als eine Darstellung in einer unformatierten Liste [Isk06, Lar07]. Durch die semantische Fixierung des Themenfeldes möglicher Suchanfragen wird zudem auch der Problematik von Synonymen begegnet. So wird bspw. eine Suche nach dem Begriff „Phyton" bei dem Suchdienst für Softwareentwickler Krugle[5] stets Ergebnisse bezogen auf die gleichnamige Programmiersprache und niemals Informationen zu einem Reptil liefern.

Bedingt durch die relative Neuheit des Feldes der spezialisierten Suche und das breite Interesse, welches diesem Thema entgegengebracht wird, existieren in einzelnen Quellen noch Unstimmigkeiten hinsichtlich der scharfen Abgrenzung einiger zentraler Begriffe. Dieses gilt bspw. für die Abgrenzung von spezialisierter Suche und dem häufiger verwendeten Begriff der vertikalen Suche. Eine Frage, die sich stellt ist, ob sich vertikale Suche lediglich auf die Spezialisierung auf bestimmte thematische Nischen beschränkt und somit nur einen Teil der spezialisier-

---

4 In gewissen Grenzen bieten generische Suchmaschinen ebenfalls diese Form der parametrisierten Suche (vgl. http://www.google.com/intl/en/help/features.html). Jedoch sind diese Funktionen vielen Nutzern nicht bekannt und werden daher auch nicht genutzt. Zudem sind sie in ihrer Handhabung umständlicher als die parametrisierte Suche spezialisierter Suchdienste, welche diese Funktion durch entsprechende grafische Elemente unterstützen.

5 http://www.krugle.com.

ten Suche abdeckt oder ob sie auch die Fokussierung bestimmter Inhaltstypen wie bspw. Musik, Quellcode, Podcasts oder Blogeinträge umfasst (MATT CUTTS in [Mac07]) und somit mit der spezialisierten Suche gleichzusetzen ist [Sul07]. In den nachfolgenden Ausführungen wird der letztgenannte Standpunkt eingenommen und die beiden Begriffe werden synonym verwendet. Wenn eine Unterteilung spezialisierter Suchdienste gefragt ist, so kann sicherlich in die Gruppe der Dienste, die sich hinsichtlich thematischer Aspekte spezialisieren und der Gruppe der Dienste, die bezüglich der Inhaltstypen fokussiert sind, unterschieden werden. Eine weitere Spezialform der spezialisierten Suche ist die lokale Suche, welche die Suchergebnisse hinsichtlich geografischer Gesichtspunkte voneinander abgrenzt. Wichtig ist zudem, dass im Folgenden lediglich die Art von Suchdiensten betrachtet wird, welche Zugriff auf Informationen gewähren, die nicht unter Kontrolle des Anbieters des Suchdienstes stehen. Aus diesem Grund fällt bspw. ein Dienst wie GoogleBase[6] oder ein Gebrauchtwagenportal wie autoscout24[7] *nicht* unter die Kategorie der hier betrachteten Suchdienste.

2 Technische Realisierungsformen spezialisierter Suche

Zur technischen Realisierung spezialisierter Suchmaschinen existieren diverse Ansätze. STEELE gibt mit seiner Taxonomie von Techniken spezialisierter Suchmaschinen (vgl. Abbildung 1) einen groben Überblick über die verschiedenen Ansatzpunkte [Ste01]. Generell existieren zwei unterschiedliche Vorgehensweise für das Design vertikaler Suchdienste. Zum einen kann ein Index Verwendung finden, welcher thematisch fokussiert ist (*Verwendung spezialisierter Indizes*), zum anderen kann der Abfrageprozess an sich auf eine spezielle Domäne hin ausgerichtet werden (*Spezialisierte Modifikation des Suchterms*). Wird keiner der beiden Ansätze implementiert, so spricht man von einer wie oben beschriebenen generischen Suchmaschine.

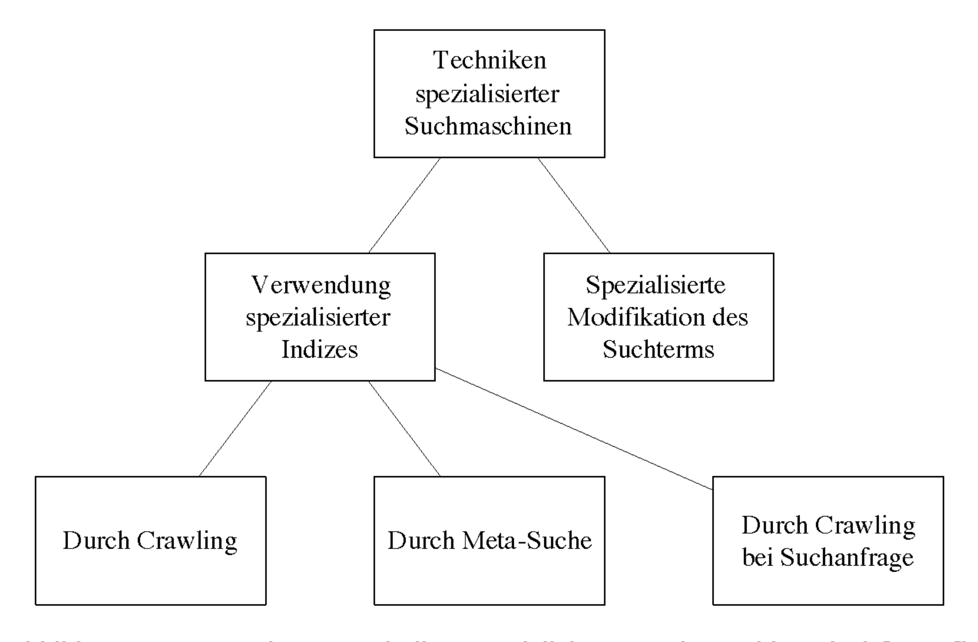


Abbildung 1: Taxonomie von Techniken spezialisierter Suchmaschinen (vgl. [Ste01])

---

6 http://www.google.com/base.

7 http://www.autoscout24.de.

Die Verwendung spezialisierter Indizes kann nach drei weiteren Ausprägungen unterschieden werden. Zum einen kann ein eigener Index aufgebaut werden, welcher lediglich thematisch relevante Informationen erfasst (*Verwendung spezialisierter Indizes durch Crawling*). Popularisiert werden diese entweder, indem lediglich die Seiten inhaltlich passender Webseiten mithilfe eines fokussierten Crawlers erfasst werden, indem die durch einen allgemeinen Crawler gelieferten Seiten in einem weiteren Schritt thematisch analysiert und gefiltert werden oder durch eine Kombination beider Ansätze. In Abhängigkeit von der Strukturiertheit der so indizierten Seiten können diese noch weiter analysiert werden und ihre Inhalte in eine streng strukturierte Form überführt werden. Die Job-Suchmaschine FlipDog[8] bspw. durchsucht die Stellenangebote von Firmenseiten und überführt diese semistrukturierten Daten in ein Standardformat, das über das Portal nach diversen Kriterien durchsucht werden kann [Ste01].

Ein weiterer Ansatz für die vertikale Suche über fokussierte Indizes verwendet keinen eigenen Index, sondern greift bei einer Metasuche auf die Indizes anderer inhaltlich relevanter Datenbanken zurück (*Verwendung spezialisierter Indizes durch Metasuche*). Diese Datenbanken sind häufig spezielle dynamische Dienste, die von den generischen Suchmaschinen nicht erfasst werden. Ein Nachteil dieses Ansatzes ist sicherlich, dass die Metasuchmaschine von den Suchfunktionalitäten der spezifischen Datenbanken abhängig ist und somit stets den kleinsten gemeinsamen Nenner aller angebundenen Diente hinsichtlich der Mächtigkeit der Suchfunktionalität darstellen kann [Ste01].

Die dritte Möglichkeit der Verwendung eines spezialisierten Indexes ist das Anstoßen eines Crawlingprozesses bei Erhalt einer Suchanfrage (*Verwendung spezialisierter Indizes durch Crawling bei Suchanfrage*). Dieser Ansatz stützt sich stark auf die Kontextinformationen einer Suche und startet vor diesem Hintergrund einen parametrisierten Crawlingprozess *nach* Eintreffen der Suchanfrage. Was für eine generische Suchmaschine nicht machbar ist, da die potenziellen Quellen relevanter Informationen zu vielfältig sind, lässt sich in spezifischen Suchsituationen durchaus realisieren. Ein Beispiel hierfür ist das Projekt „AHOY“[9], welches einen vertikalen Suchdienst zur Auffindung der Homepages natürlicher Personen entwickelt hat. Anhand persönlicher Daten, wie Vor- und Nachnamen einer Person, der E-Mail-Adresse, der Firma und dem Heimatland, versucht der Dienst, mithilfe einer Heuristik die entsprechende Seite aufzuspüren [Ste01].

Für die zweite große Gruppe der technischen Realisierungsmöglichkeiten spezialisierter Suchdienste wird die vertikale Fokussierung über die domänenspezifische Modifikation der Suchterme – transparent für den Nutzer – erreicht. Über die Integration intelligenter Klassifikationsalgorithmen – wie bspw. Support Vector Machines – in die Suchmaschinen können diese darauf trainiert werden, Suchanfragen für eine spezielle Domäne zu optimieren. Die so modifizierten Suchanfragen werden an generische Suchmaschinen weitergeleitet und von diesen beantwortet [Ste01].

---

8 http://www.flipdog.com.

9 http://www.cs.washington.edu/research/ahoy.

3 Die ökonomische Perspektive spezialisierten Suchens

Google hat gezeigt, dass die Verknüpfung von Suchanfragen und Werbeeinblendungen einen großen und lukrativen Markt bildet. Daher investierten seit Anfang 2004 Venture Capital-Geber fast 350 Mio. $ in Start-up-Unternehmen, die sich der Internetsuche verschrieben haben. Der überwiegende Teil dieser Unternehmen nimmt jedoch nicht den direkten Konkurrenzkampf mit dem unangefochtenen Platzhirschen Google auf, sondern bietet spezialisierte Suchdienste an [Hel07].

Nach der Meinung vieler Experten stellen diese spezialisierten Suchdienste nicht lediglich eine gegenwärtige Modeerscheinung dar, sondern werden in Zukunft einen nicht unerheblichen Anteil am weltweiten Markt der Suchdienste für sich beanspruchen [Per06, Com05]. Hier wird eine ähnliche Entwicklung wie im amerikanischen Markt für Kabel-TV prognostiziert. In dieser Domäne konnten seit Ende der achtziger Jahre Nischensender mit thematisch fokussierten Programmen den großen, breit aufgestellten Sendernetzwerken wie ABC, CBS und NBC in erheblichem Maße Marktanteile abnehmen [Lar07, Per06].

Dementsprechend wird sich ebenfalls der Anteil an Werbebudgets, die für Onlinewerbung verausgabt werden, zugunsten der spezialisierten Suchmaschinen verschieben. Jedoch sehen die Analysten diesen Trend nicht gleichermaßen für alle Domänen spezialisierter Suche. Der Großteil der heute schon durch vertikale Suchdienste vereinnahmten Werbeausgaben verteilt sich auf lediglich wenige vertikale Geschäftsfelder. So wurden 79 % der 2004 für bezahlte Suche generierten Umsätze von 2,6 Mrd. $ auf die vier Domänen Handel (32 %), Finanzdienstleistung (21 %), Unterhaltungsindustrie (16 %) und Reisen (10 %) verteilt [Res05, Kuc05]. In dieser Konstellation hat jedoch vor allem der Onlinemarkt für Reisen in den letzten drei Jahren überdurchschnittlich an Größe zugenommen. So wuchs diese Industrie in Europa 2005 um 49 % und soll nach Schätzungen mit dieser Wachstumsrate bis 2009 konstant an Größe zulegen [Mit07b].

Doch nicht nur aus Sicht der Betreiber von Suchmaschinen als Anbieter von Werbeflächen ist der Markt für spezialisierte Suchdienste interessant. Auch der Gruppe der Werbenden, die für die Schaltung ihrer Werbung eine geeignete Plattform suchen, bieten spezialisierte Suchdienste zwei entscheidende Vorteile. Zum einen sind die Konditionen für die Werbung auf spezialisierten Suchmaschinen meistens wesentlich günstiger als auf den großen generischen Suchmaschinen wie Yahoo! und Google [Sch07]. Zum anderen bietet die thematische Spezialisierung vertikaler Suchmaschinen den Vorteil, dass das Publikum, welches auf spezialisierte Suchmaschinen gelockt wird, hinsichtlich seiner Intentionen wesentlich besser eingeschätzt werden kann. „Advertising on vertical search engines is all about fishing where the fish are“ (SCOTT VIRKLER in [Per06]). Durch den engeren thematischen Zusammenhang zwischen Informationsbedürfnis des Nutzers und angezeigter Werbung steigt die Rate der tatsächlich verfolgten Werbelinks [Kuc05]. Dieses rechnet sich vor allem für Preismodelle, bei denen der Werbende schon allein für die Anzeige seiner Werbung zahlt. Zudem sind die Nutzer spezialisierter Suchmaschinen wesentlich zielgerichteter und häufig wesentlich näher an einer Kaufentscheidung als Kunden generischer Suchmaschinen. So wird der Nutzer eines Gebrauchtwagenportals sicherlich mit einer höheren Wahrscheinlichkeit ein Kraftfahrzeug erwerben wollen als der Nutzer einer generischen Suchmaschine, welcher nach dem gleichen Fahrzeugtyp sucht. Dementsprechend wird auch der Klick auf eine dort angezeigte Werbung für Autoversicherungen „ernster gemeint“ sein als im Falle der Allzwecksuchmaschine. Dies ist vor allem dann für den Werbenden interessant, wenn Werbekosten proportional zu der Häufigkeit der Klicks auf angezeigte Werbung

entstehen. Die Kombination all dieser Faktoren verspricht dem Werbenden daher einen wesentlich höheren Return on Investment, falls er sich bei der Schaltung von Onlinewerbung für eine vertikalen Suchmaschine und gegen eine generische entscheidet [Sch07].

Da die großen drei Allzwecksuchmaschinen Google, Yahoo! und MSN jedoch zurzeit noch die überwältigende Mehrheit an Suchanfragen bedienen, müssen bei der Entscheidung für die Schaltung von Werbung auf spezialisierten Suchdiensten diese mit Bedacht ausgewählt werden. Andernfalls ist der über diese Seiten erzeugte zusätzliche Verkehr auf der eigenen Seite zu gering, als dass sich die Werbung lohnen würde. Zur Entscheidungsunterstützung empfiehlt SCHIFF den intensiven Einsatz von Webanalysewerkzeugen um hierüber den werbegenerierten Publikumsverkehr und die hierüber erzeugten Umsätze – im Falle eines Onlineshops – messen und den einzelnen Werbeträger zurechnen zu können. In einem Trail-and-Error-Verfahren muss jeder Werbende herausfinden, welcher spezialisierte Suchdienst für ihn das beste Kosten-Nutzen-Verhältnis erbringt [Sch07].

Offen bleibt jedoch die Frage, inwieweit diese Entwicklung zu einer größeren Vielfalt an Suchdiensteanbietern führt. So offeriert das Angebot eines spezialisierten Suchdienstes gegenwärtig wohl eine probate Markteinstiegsstrategie für kleine Unternehmen [Sok06], doch sind auch die etablierten Suchdiensteanbieter nicht untätig auf dem Gebiet der spezialisierten Suche [Kop05]. Daher wird sich zeigen, ob auch auf Dauer lukrative Nischen im Suchmarkt von den großen Anbietern unbesetzt bleiben [Isk07a]. Daher sollten spezialisierte Neueinsteiger in den Markt der Suchmaschinen frühzeitig versuchen, sich auf ihrem Gebiet als Marke zu etablieren, um hierüber bei einem verstärkten Engagement der großen Anbieter im Bewusstsein der potenziellen Nutzer zu verbleiben [Isk07b].

4 Soziale Ansätze spezialisierter Suche

Im Zuge der Web 2.0-Bewegung hat sich das Web weg von einer Quelle statischer Inhalte hin zu einer Plattform kollaborativer, selbstorganisierter Produktion diverser Inhaltsformen entwickelt, in der die Nutzer des Webs Inhalt nicht nur konsumieren, sondern gleichfalls aktiv an deren Gestaltung mitwirken. Jeder Einzelne kann zur Bildung dieser Weisheit der Massen (*Wisdom of the Crowds*) in Blogs, Wikis, Foren, Fotogalerien und auf Videoportalen beitragen [TW06]. Diese neuartigen Informationsformen und die Charakteristika ihrer Entstehungsprozesse bieten für die Domäne der spezialisierten Suchdienste nicht nur weitere Spezialisierungsfelder, sondern ebenfalls verschiedene Ansätze, spezialisierte Suchvorgänge aktiv zu gestalten.

Bei spezialisierten Suchmaschinen kann die Auswahl eines Nutzers aus gelieferten Suchergebnissen dazu genutzt werden, die Relevanz der einzelnen Ergebnisse zu bewerten. Über die Erfassung der händischen Auswahl relevanter Informationen durch die Nutzer eines Suchdienstes kann bei nachfolgenden Anfragen zum gleichen Suchterm die Reihenfolge der präsentierten Einträge angepasst werden [Mit07a]. Derartige manuell erzeugte Relevanzindikatoren können automatisierte Bewertungsverfahren wie den Page-Rank-Algorithmus, der auf rein strukturellen Informationen operiert, sinnvoll ergänzen. Einen entsprechenden Ansatz verfolgt das Projekt Wikia Search[10], welches die Entwicklung einer Plattform für die Gestaltung freier Suchmaschi-

10 search.wikia.com.

nen vorantreibt, die sich dadurch auszeichnen, dass Suchergebnisse von den Nutzern editiert und anderen zur Verfügung gestellt werden können. Im Einklang mit dem „Wisdom of the Crowds"-Prinzip sollen hier die Nutzer der Suchmaschinen aktiv, durch Bewertung der Suchergebnisse, zu einem hohen Standard an Ergebnisqualität beitragen [Dor06]. Diese Technik ist natürlich sowohl auf generische als auch auf spezialisierte Suchmaschinen anwendbar, wobei bei letzteren die Bewertung hinsichtlich der Relevanz zu einem bestimmten Themenbereich sicherlich einfacher zu bewerkstelligen ist, als bei dem Ergebnis einer generischen Suchmaschine kontextfrei die Qualität eines Links zu bemessen.

Ein Problem, das sich beim Bau spezialisierter Suchdienste stellt, ist die Auswahl der potenziell relevanten Quellen. Hier können die im Zuge der Web 2.0-Bewegung eine weite Verbreitung gefundenen sozialen Lesezeichendienste und Verschlagwortung – neudeutsch auch als Tagging bezeichnet – Hilfestellung leisten. Quellen, die mit demselben Tag versehen sind oder in verschiedenen Lesezeichensammlungen unter demselben Ordner zusammengefasst sind, können auf entsprechend thematisch zusammenhängende Quellen im Web hinweisen. Derartige Analysen können als Ausgangspunkt für die automatische Identifikation thematisch verwandter Teilbereiche des Internets dienen. Hierüber kann die Eingrenzung der Quellen spezialisierter Suchmaschinen effektiver und effizienter gestaltet werden.

Eine weitere Inkorporation sozialer Aspekte in vertikale Suchen stellen Dienste wie Google Coop oder Yahoo! Pipes dar (vgl. Abschnitt 5.3). Hier können sich Nutzer nicht nur individuelle Suchmaschinen zusammenstellen, sondern andere Nutzer können diese ebenfalls nutzen und weiterverwenden.

Einen weiteren Versuch, die *Weisheit der Massen* für sich zu nutzen, unternehmen die großen Anbieter von Suchdiensten, indem sie die APIs ihrer Services offen legen und jeden ermutigen, so genannte Mash-Ups zu erzeugen. Als Mash-Up bezeichnet man die Kombination bzw. Erweiterung mehrere (Web)Services, um hierüber einen neuen Service zu erzeugen. Auf diese Weise verschaffen sich große Anbieter von (vertikalen) Suchdiensten wie Google, Yahoo! oder Amazon kostengünstig Zugriff auf externe Ideen, Talente und innovative Ressourcen, welche nicht selten in die Innovationsstrategien dieser Unternehmungen einfließen [TW06].

5 Anbieter spezialisierter Suchdienste

Es existiert neben den allseits bekannten generischen Suchmaschinen wie Google, Yahoo! und Altavista eine immense Fülle an weniger bekannten vertikalen Suchmaschinen. Dieses Angebot ist mittlerweile so unüberschaubar geworden, dass auch Kataloge und Suchmaschinen für Suchmaschinen – die eigentlichen *Metasuchmaschinen* – wie der SearchEngineGuide[11] im World Wide Web ihre Existenzberechtigung besitzen. Aus diesem Grund sind die nachfolgenden Ausführungen weit davon entfernt, einen umfassenden Überblick über den Markt der vertikalen Suchdiensteanbieter zu geben. Es werden lediglich einige Beispiele angeführt.

---

[11] http://www.searchengineguide.com/searchengines.html. Vgl. auch *Specialized Search Engines and Directories* unter http://webquest.sdsu.edu/searching/specialized.html.

Untergliedern lassen sich die vertikalen Suchdienste auf der einen Seite hinsichtlich ihres Anbieters. So gibt es einerseits die Suchdienste autonomer Nischenanbieter (vgl. Abschnitt 5.1) und andererseits bieten auch die großen Anbieter horizontaler Suchmaschinen vertikale Zusatzdienste an (vgl. Abschnitt 5.2). Zudem existiert eine dritte Kategorie von Diensten, die sich dadurch auszeichnen, dass die Spezialisierung von jedem durchgeführt und als Service genutzt werden kann (vgl. Abschnitt 5.3).

5.1 Spezialisierte Suche von Nischenanbietern

Vertikale Suchmaschinen von Nischenanbietern existieren in den verschiedensten Bereichen. So bieten Dienste wie KidsClick![12] oder AskForKids[13] eine Art horizontale Suchmaschine, vertikal fokussiert auf Inhalt, die auch für Kinder bestimmt sind. Die Suche nach dem Begriff „porn" bei dem erstgenannten Anbieter führt bspw. zu keinem Ergebnis. Ebenso existieren diverse Suchdienste zum Aufspüren von Biografien, Bildern, Zeitungseinträgen oder lizenzfreier Literatur [Dod06].

Ein Anbieter mit einem der größten Angebote mit ca. 180 vertikalen Suchmaschinen ist LookSmart[14] [Ste05]. Der Suchdienst Technorati[15] ist hingegen ein Dienst, welcher es geschafft hat, auf dem Gebiet der Blogsuche bereits eine Art Marke zu etablieren. Dieses räumt ihm gute Chancen ein, auch im Konkurrenzkampf mit den Großen der Branche überleben zu können [Isk07b]. Auf dem Gebiet der Onlinereisesuchmaschinen ist Expedia[16] der unangefochten größte Anbieter im Markt. In der Rangliste der Webseiten, die den meisten Verkehr auf sich ziehen, rangiert Expedia auf Rang 409 und erzielte so im Jahr 2006 einen Erlös von 2,2 Mrd. $ [Mit07b]. Auf dem Gebiet der vertikalen Shoppingunterstützung ist become.com[17] sicherlich zu nennen. Dieser Dienst basiert im Gegensatz zu vielen anderen Preisvergleichsdiensten, welche nach dem Push-Prinzip mit Preisinformationen von den Produktanbietern versorgt werden, auf einem crawlerbasierten Pull-Prinzip. Darüber hinaus bietet der Dienst die Möglichkeit, Informationen zu einem Produkt von Interesse im Web zu lokalisieren. Diese Informationen gehen über die Bewertungen, welche bei vielen anderen Shoppingportalen auf der Seite selbst gespeichert werden, hinaus.

5.2 Spezialisierte Suche von etablieren Anbietern generischer Suche

Vor dem Hintergrund des zunehmenden Engagements der großen Anbieter horizontaler Suchmaschinen auf dem Gebiet der vertikalen Suche und dem hier zu erwartenden Konkurrenzkampf mit autonomen Nischenanbietern, ist die Frage nach den spezialisierten Angeboten der Großen von besonderem Interesse. Bei einem Blick auf das Portfolio der Nummer eins auf dem Markt der Suchmaschinen zeigt sich, dass Google vornehmlich vertikale Suchen auf bestimmte Informationstypen anbietet. So bietet Google spezialisierte Suchdienste für Blogeinträge[18], Bü-

---

12 http://www.kidsclick.org.
13 http://www.askforkids.com.
14 http://search.looksmart.com.
15 http://technorati.com.
16 http://www.expedia.com.
17 http://www.become.com.
18 http://blogsearch.google.com.

cher[19], Open Source Code[20], Börsenkurse[21], Bilder[22], Nachrichten[23], Patente[24], Produktangebote in Onlineshops[25], wissenschaftliche Literatur[26] sowie Videos[27] an [Sul07]. Allerdings lassen sich auch ein paar rein thematisch fokussierte Suchmaschinen von Google finden. So werden vertikale Suchdienste zu den Themen US Regierung[28], Linux[29], BSD[30], Apple Macintosh[31] und Microsoft[32] von Google offeriert. Für Yahoo! stellt sich ein sehr ähnliches Bild dar. Erwähnenswert ist hier lediglich, dass Yahoo! eine Form der lokalen Suche anbietet, über die ein Nutzer nach Immobilien[33] anhand geografischer Parameter suchen kann.

5.3 Spezialisierte Suche als Service

Neben den diversen Angeboten vertikaler Suchdienste durch Anbieter wie Technorati, become.com, Google oder Yahoo!, existieren Dienste im Netz, die es jedem erlauben, seine eigene vertikale Suche zu implementieren. Mit Google Custom Search Engine[34] bietet der Branchenprimus die Möglichkeit, eine spezialisierte Suchmaschine zu realisieren, welche eine vordefinierte Menge an Webadressen durchsucht [Mac07]. Die Technologie bedient sich hierbei dem Index Googles generischer Suchmaschine. Hierüber können bspw. sehr leicht Suchfunktionalitäten in einzelne Homepages als Service integriert werden. Über sein Portal Google Co-op[35] wird zusätzlich die Chance geboten, seine erstellten vertikalen Suchmaschinen anderen zur Verfügung zu stellen [Arr06].

In einem etwas anderen Ansatz bietet auch Yahoo! kostenlos die Möglichkeit, individuelle vertikale Suchmaschinen zu erzeugen. Mit Yahoo! Pipes[36] wird ein Werkzeug zur Verfügung gestellt, welches es komfortabel über eine grafische UI erlaubt, die unterschiedlichsten Datenquellen zu integrieren, zu filtern und interaktiv abzufragen. Mögliche Datenquellen sind Newsfeeds, soziale Bookmarking-Dienste, Blogs, Bildportale, lokale Suchdienste, Musikportale etc. Mit Blick auf die Taxonomie von Techniken spezialisierter Suchmaschinen (vgl. Abschnitt 2) fallen die über Yahoo! Pipes realisierten Suchdienste unter die Kategorie *Verwendung spezialisierter Indizes durch Metasuche*. Der soziale Aspekt ist zudem ein zentraler Gedanke dieses Dienstes. Pipes anderer Nutzer können verwendet werden und eigenen Pipes als Datenquelle dienen.

---

[19] http://books.google.com.
[20] http://www.google.com/codesearch.
[21] http://finance.google.com/finance.
[22] http://images.google.com.
[23] http://news.google.com.
[24] http://www.google.com/patents.
[25] http://www.google.com/products.
[26] http://scholar.google.com.
[27] http://video.google.com.
[28] http://www.google.com/unclesam.
[29] http://www.google.com/linux.
[30] http://www.google.com/bsd.
[31] http://www.google.com/mac.html.
[32] http://www.google.com/microsoft.html.
[33] http://realestate.yahoo.com.
[34] http://www.google.com/coop/cse.
[35] http://www.google.com/coop.
[36] http://pipes.yahoo.com.

6 Abschließende Betrachtung

Ziel dieses Beitrags war es, einen möglichst umfassenden Überblick über das Thema der spezialisierten Suche zu geben. Chancen und Restriktionen dieses Konzeptes wurden sowohl aus technischer, aus ökonomischer als auch aus sozialer Perspektive beleuchtet. Festzustellen ist, dass sich das Konzept der vertikalen Suche sowohl aufgrund seiner ökonomischen Vorteile als auch bedingt durch seine überlegene Effektivität in Zukunft einen festen Platz in der Domäne der Internetsuche sichern wird. Unabhängig jedoch davon, wie erfolgreich spezialisierte Suchdienste am Ende sein werden, wird es stets einen Markt für generische Suchmaschinen geben. Es ist ökonomisch nicht sinnvoll, für jede erdenkliche thematische Nische spezialisierte Suchmaschinen anzubieten [Isk06]. Zudem wird es auch in Zukunft einen nicht unerheblichen Anteil an Nutzern von Suchmaschinen geben, die nicht gewillt sind, den jeweiligen Suchkontext näher zu spezifizieren – sei es durch die Auswahl einer passenden spezialisierten Suchmaschine oder durch Konfiguration einer generischen Suchmaschine. Diese Nutzer präferieren es, einen zentralen Einstiegspunkt mit einer übersichtlich gestalteten Oberflächen für ihre Suchanfragen zu haben und sind zufrieden mit dem thematisch weit gefächerten Überblick, den die Suchergebnisse generischer Suchmaschinen widerspiegeln. Sie ziehen es vor, Suchergebnisse schnell geliefert zu bekommen und diese anschließend händisch filtern zu müssen, anstatt im Vorfeld des Suchprozesses umfangreiche Eingaben tätigen zu müssen [Mit07a]. Eine mögliche Verbesserung der Suchbedingungen für diese Nutzergruppe von elektronischen Suchdiensten stellt der relativ neue Ansatz der universellen Suche dar. Hier wird ausgehend von einem einfachen, generischen Suchinterface eine Vielzahl spezialisierter Suchmaschinen angesprochen. Die Suchergebnisse werden nach Abschluss des Suchprozesses integriert und unter Explikation des jeweiligen Kontextes der Ergebnisse dem Nutzer präsentiert [Sul07, Cat07].

Literaturverzeichnis

[Arr06] Michael Arrington. Google co-op launches [online]. Oct 2006. Available from: http://www.techcrunch.com/2006/10/23/google-custom-search-tomorrow/.

[Bat05] John Battelle. The Search – How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture. Portfolio, 2005.

[Bro02] Andrei Broder. A taxonomy of web search. SIGIR Forum, 36(2):3-10, 2002.

[Cat07] Josh Catone. How does Google Universal Search compare to Yahoo! , Alpha, Ask X, Microsof Imagine Live and Google's own Searchmash? [online]. 5 2007. Available from: http://www.readwriteweb.com/archives/how_does_google_universal_search_c%ompare.php.

[Com05] Devin Comiskey. Vertical search engines look to connect buyers and sellers [online]. 10 2005. Available from: http://www.ecommerce-guide.com/solutions/affiliate/article.php/3559691.

[Dod06] Bernie Dodge. Specialized search engines and directories [online]. 12 2006. Available from: http://webquest.sdsu.edu/searching/specialized.html.

[Dor06] James Doran. Founder of wikipedia plans search engine to rival google [online]. 12 2006. Available from: http://technology.timesonline.co.uk/tol/news/tech_and_web/article126411 %7.ece.

[Hel07] Miguel Helft. In silicon valley, the race is on to trump google [online]. 1 2007. Available from: http://www.nytimes.com/2007/01/01/technology/01search.html?ex=132530760 %0&en=5fb9c9a36e8ccb8d&ei=5088&partner=rssnyt&emc=rss.

[Isk06] Alex Iskold. Watch out google, vertical search is ramping up! [online]. 9 2006. Available from: http://www.readwriteweb.com/archives/vertical_search.php.

[Isk07a] Alex Iskold. Google universal search – is vertical search space finished? [online]. 5 2007. Available from: http://www.readwriteweb.com/archives/google_universal_search_vertical_s %earch_finished.php.

[Isk07b] Alex Iskold. The race to beat google [online]. 1 2007. Available from: http://www.readwriteweb.com/archives/the_race_to_beat_google.php.

[Kar07] Nitin Karandikar. Top 17 search innovations outside of google [online]. 5 2007. Available from: http://www.readwriteweb.com/archives/top_17_search_innovations.php.

[Kop05] Verne Kopytoff. New search engines narrowing their focus [online]. 4 2005. Available from: http://sfgate.com/cgi-bin/article.cgi?file=/c/a/2005/04/04/BUGJ9C20VU1.%DTL.

[Kuc05] Susan Kuchinskas. Doing a vertical search [online]. 8 2005. Available from: http://www.internetnews.com/xSP/article.php/3526391.

[Lar07] Steve Larsen. Google: The next CBS? [online]. 7 2007. Available from: http://altsearchengines.com/2007/07/26/google-the-next-cbs/.

[Mac07] Richard MacManus. Interview with google's matt cutts about next-generation search [online]. 3 2007. Available from: http://www.readwriteweb.com/archives/interview_with_matt_cutts_next_gen %eration_search.php.

[Mit07a] Andy Mitchell. Meta vertical search by 3 dogs bark! [online]. 8 2007. Available from: http://altsearchengines.com/2007/08/03/meta-vertical-search-by-3-dogs-b %ark/.

[Mit07b] Sramana Mitra. Travel and web 3.0: Overview [online]. 4 2007. Available from: http://sramanamitra.com/2007/04/23/travel-web-30-overview/.

[Per06] Juan Carlos Perez. What is vertical search? [online]. 1 2006. Available from: http://www.infoworld.com/article/06/01/18/74292_HNverticalsearch_1.html %.

[Res05] Jupiter Research. Jupiter research projects vertical search will drive indstry growth, with four industries dominating search spending [online]. 2 2005. Available from: http://www.jupitermedia.com/corporate/releases/05.02.23-newjupresearch.%html.

[Rij79] CJ Rijsbergen. Information Retrieval. Butterworths, London, 1979. Available from: http://sfx.hbz-nrw.de/sfx_ulbms?sid=google;auinit=CJ;aulast=Rijsbergen; %title=Information %20retrieval;genre=book;isbn=0408707178;date=1975.

[RL04] Daniel E. Rose and Danny Levinson. Understanding user goals in web search. In WWW '04: Proceedings of the 13th international conference on World Wide Web, pages 13-19, New York, NY, USA, 2004. ACM Press.

[Sch07] Jennifer Schiff. Does it pay to advertise on vertical search engines? [online]. 4 2007. Available from: http://www.ecommerce-guide.com/solutions/article.php/3669361.

[Sok06] Emre Sokullu. Search 2.0 – what's next? [online]. Dec 2006. Available from: http://www.readwriteweb.com/archives/search_20_what_is_next.php.

[Ste01] R Steele. Techniques for specialized search engines. In Proceedings of the International Conference on Internet Computing '01, pages 25-28, Las Vegas, 2001.

[Ste05] Greg Sterling. Looksmart goes deep vertical [online]. 10 2005. Available from: http://www.searchenginejournal.com/looksmart-goes-deep-vertical/2409/.

[Sul07] Danny Sullivan. Google 2.0: Google universal search [online]. 5 2007. Available from: http://searchengineland.com/070516-143312.php.

[TW06] Don Tapscott and Anthony D. Williams, Wikinomics – How Mass Collaboration Changes Everything, New York 2006.

Personalisierte Suche

Milan Karow

Abstract: Seit dem Siegeszug des PageRank-Alghorithmus und der damit einhergehenden Marktdominanz der Suchmaschine Google stoßen universelle Suchmechanismen in ihrer Leistungsfähigkeit an semantische Grenzen. Der anonyme Transaktionscharakter der Suchanfragen bewirkt gleiche Ergebnisse bei gleichen Anfrageausdrücken – für bessere Suchergebnisse benötigt die Suchmaschine mehr Informationen. Personalisiertes Suchen soll dabei helfen, für einen individuellen Nutzer potenziell interessante Inhalte bevorzugt zu präsentieren. Der vorliegende Beitrag gibt eine kurze Einführung in die Hintergründe, Prinzipien und potenzielle Gefahren dieser Technologie.

1 Suchen im World Wide Web

1.1 Suchmaschinen als Instrument des Information Retrieval

Suchmaschinen sind ein zentrales Instrument des *Information Retrieval*. Sie dienen zur Wiederauffindung von Informationen in Computersystemen. Ziel ist dabei die Minimierung des zur Suche benötigten Aufwands.

Das World Wide Web (WWW) als öffentliches, höchst dynamisches und heterogenes Netzwerk von Dokumenten auf vernetzten Rechnersystemen konfrontiert den Nutzer mit einer Datenflut, die ohne strukturierende Automatismen kaum nutzbar wäre. Relevante, dem Nutzer jedoch bislang nicht bekannte Informationen sind dabei manuell nicht mit vertretbarem Aufwand zu ermitteln [8]. Ein Projekt der *Induction of Linguistic Knowledge Research Group* an der Universität von Tilburg schätzt die Größe des WWW auf mindestens 23 Mrd. Dokumente (Stand: Ende August 2007). Die Schätzung bezieht sich jedoch auf die vereinigte Indexmenge der Suchmaschinen Yahoo! [→b], Google [→a], MSN Live Search [→c] und Ask [→d] – sodass von einer deutlich größeren realen Anzahl an Dokumenten im WWW ausgegangen werden kann [5]. Suchmaschinen bieten diese strukturierende Funktionalität, indem sie vollautomatisch Dokumente indizieren und Linkstrukturen auswerten.

1.2 Allgemeine Funktionsprinzipien von Suchmaschinen

Die Aufgabe von Suchmaschinen ist zunächst die Steigerung der *Effizienz* und *Effektivität* von Retrievalaufgaben im WWW. KOBAYASHI und TAKEDA stellen als Metrik für Suchmaschinen einführend die Konzepte *precision* und *recall* vor [8]. Die *precision* ergibt sich dabei als Verhältnis von den relevanten Dokumenten zu allen gefundenen Dokumenten einer Suchanfrage:

$$\text{precision} = \frac{\text{number of relevant documents}}{\text{number of retrieved documents}}$$

Unter *recall* wird das Verhältnis zwischen den gefundenen relevanten Dokumenten einer Suchanfrage und allen für den Nutzer relevanten Dokumenten verstanden:

$$\text{recall} = \frac{\text{number of relevant, retrieved documents}}{\text{total number of relevant documents}}$$

Die Autoren räumen ein, dass der tatsächliche *recall* einer Suchanfrage kaum messbar ist, da die Gesamtheit der relativen Dokumente nicht zu ermitteln sei. Auch ist die Präzision der Suche nicht das vordergründige Gütekriterium in der Praxis. Solange die potenziell relevanten Ergebnisse innerhalb der Trefferliste schnell auffindbar sind, ist die totale Treffermenge (und damit die Menge nicht-relevanter Treffer im Suchergebnis) irrelevant. Demzufolge ist die *Anzeigereihenfolge* der Trefferliste entsprechend der Relevanz der gefundenen Dokumente ein entscheidender Faktor.

Viele Forschungsarbeiten zur Qualitätssteigerung der Websuche beschäftigten sich in den letzen Jahren mit der Herleitung eines Sortieralgorithmus, welcher hinsichtlich verschiedener Kriterien optimierte Ergebnislisten erzeugen soll.[1] Frühe Rankings waren dabei unabhängig von der Suchanfrage und sortieren die indizierten Dokumente statisch nach einem bestimmten Qualitätskriterium. Das prominenteste Beispiel für einen solchen Algorithmus ist unbestreitbar der *PageRank* [10], welcher den Grundstein des Markterfolgs des Suchmaschinenanbieters Google legte. Das Prinzip dabei ist einfach: Ein Dokument im WWW wird danach gewichtet, wie viele andere Dokumente darauf verweisen. Diese Verweise haben einen umso höheren Einfluss, je höher das Gewicht des verweisenden Dokumentes ist. Formal definiert sich der PageRank wie folgt:

$$PR_i = \frac{i-d}{N} + d \sum_{\forall\{(j,i)\}} \frac{PR_j}{C_j}$$

Dabei steht N für die Gesamtzahl der indizierten Seiten, d stellt einen Dämpfungsfaktor zwischen 0 und 1 dar und C_j die Anzahl der insgesamt verlinkten Dokumente, unter denen der Rang aufgeteilt wird.

1.3 Herausforderungen des Marktes

Die Verwendung statischer Ranking-Mechanismen wie dem *PageRank* hat in den vergangenen Jahren zum Phänomen der so genannten *Search Engine Optimization (SEO)* geführt.[16] Dabei handelt es sich um eine Form des Onlinemarketing, welche gezielt die Ranking-Mechanismen von Suchmaschinen zu unterminieren versucht. Das Ziel ist dabei, die eigene Webseite möglichst weit oben in möglichst vielen Trefferlisten zu positionieren. Für Suchmaschinenbetreiber ist diese Art der Werbung mehrfach unerwünscht: Zunächst verschlechtern solche Manipulationen die Qualität der Suche, da dem Nutzer unerwünschte kommerzielle Angebote innerhalb der Treffer präsentiert werden. Dies hat entsprechend negative Einflüsse auf die Glaubwürdigkeit der betroffenen Suchportale [14]. Des Weiteren stellt diese Art der Onlinewerbung eine direkte

---

1 Eine Übersicht findet sich bei [1]

Konkurrenz zu den bezahlten Werbeflächen auf den Portalseiten der Suchmaschinenanbieter dar und schadet damit dem Geschäftsmodell der Betreiber [16].

Die Suchmaschinenbetreiber sind durch diese Entwicklungen ständig gezwungen, ihre Suchmechanismen zu verfeinern, um entsprechende Exploits unschädlich zu machen. Aktuelle Ranking-Algorithmen berücksichtigen daher eine Vielzahl unterschiedlicher Informationen zur Berechnung ihrer Rankings. Verständlicherweise sind diese Algorithmen nicht öffentlich. Auf der Gegenseite ist jedoch ein Markt für SEO-spezialisierte Firmen entstanden, die ihren Kunden gegen Bezahlung möglichst hohe Ranking-Plätze verschaffen [16].

Eine wichtige Herausforderung des Marktes ist, dass Nutzer von Suchmaschinen in der Regel keine Experten auf dem Gebiet des Information Retrieval sind. Entsprechend werden syntaktische Möglichkeiten der Suchmaschinen nicht ausgenutzt, Suchbegriffe sind häufig kurz und mehrdeutig. Ein Nutzerkontext (z. B. durch Logging) könnte hier – ohne dem Nutzer explizites Zutun abzufordern – zusätzliche, den Anfragestring näher beschreibende Informationen bereitstellen.

2 Motivation für die Personalisierung der Websuche

2.1 Motivation aus Anbietersicht

Kommerzielle Anbieter von Suchmaschinen finanzieren sich ohne Ausnahme durch Werbung, welche entweder auf dem Suchportal direkt oder aber in den jeweiligen Ergebnislisten der Suchanfragen platziert ist. Der Wert der platzierten Werbung leitet sich zunächst von der Popularität der Seite ab, in welche der jeweilige Werbeträger eingebettet ist. Als erste essenzielle Motivation der Suchmaschinenbetreiber ist also die Steigerung der Popularität des Suchportals zu sehen. Die personalisierte Suche ist dabei eine qualitätssteigernde Maßnahme, die der Diversifizierung des angebotenen Dienstes dient [14].

Darüber hinaus ist die Sammlung nutzerbezogener Daten für den Suchmaschinenbetreiber interessant, um kommerzielle Werbeangebote kontextbezogen platzieren zu können. Bereits heute werden Werbeanzeigen auf den führenden Suchportalen abhängig von der jeweiligen Anfrage (z. B. als so genannte *sponsored links*) platziert. Diese Platzierungen sind jedoch statisch und nutzerunabhängig. Präzisere Nutzerprofile könnten den Erfolg solcher Werbeformen verbessern, indem gezielt auf das Suchverhalten des jeweiligen Nutzers reagiert wird. So könnte bspw. anhand der Suchbegriffe und der angeklickten Treffer eine Nutzerklassifikation stattfinden, um so passende Werbeangebote anbieten zu können, die nicht nur die jeweils letzte Suchanfrage berücksichtigen.

Als Beispiel: Ein Nutzer, welcher in seiner Suchhistorie nach „free cell phone games“ und „ringtones“ und vermehrt Adressen einer bestimmten Stadt gesucht hat, könnte ggf. aufgeschlossener gegenüber Werbung von lokalen Anbietern für Mobilfunkzubehör sein.

2.2 Motivation aus Nutzersicht

Nutzer von Suchportalen verfolgen in erster Linie ihre Retrieval-Ziele und sind daher interessiert an einem möglichst präzisen Ergebnis mit möglichst geringem Aufwand. Dabei gehen unterschiedliche Nutzer jedoch heterogen vor [15]. WHITE und DRUCKER identifizieren dabei zwei Hauptklassen von Nutzern: *Navigatoren* und *Kundschafter*.

Die Gruppe der *Navigatoren* klassifiziert Nutzer, deren Suchverhalten stark strukturiert ist. Diese Nutzergruppe weist sich dadurch aus, sehr zielorientiert eine bestimmte Information anzusteuern. Dabei wird das jeweilige Suchproblem sequenziell angegangen, bestimmte Domains mit empfunden hoher Relevanz werden mehrfach besucht. Die Gruppe der *Kundschafter* zeichnet sich durch eine variable, chaotisch anmutende Suchstrategie aus, mit vielen Verzweigungen, vielen Suchanfragen pro Sitzung und Besuch vieler neuer Domains. Die Autoren räumen ein, dass diese Archetypen in der Praxis eher als Mischform auftreten und darüber hinaus vom jeweiligen Suchproblem abhängen [15].

Personalisierte Suchmechanismen können beide Suchmuster unterstützen. Die Definition eigener Suchmaschinen (z. B. Eurekster, Rollyo), welche bestimmte Webseiten bevorzugt durchsuchen, kommt dem navigierenden Suchen entgegen. Ebenso sorgen Suchprotokolle (z. B. Google) für einen schnellen erneuten Zugang zu bereits gefundenen Ergebnissen. Toolgestützte Bookmarking-Mechanismen (z. B. deli.cio.us) helfen, bei explorativen Suchanfragen den Überblick zu behalten, Social Bookmarking-Angebote ermöglichen einen breiten Zugang zu Suchthemen.

3 Prinzipien und Technologie

3.1 Personalisierung durch Logging

Personalisiertes Suchen im Rahmen des Web 2.0 lässt sich in zwei grundlegende Kategorien bzgl. der jeweils zugrunde liegenden Prinzipien klassifizieren. Bei der *personalisierten Suche durch Logging* handelt es sich um eine Suchtechnologie, welche das Verhalten eines Nutzers auf einem Suchportal aufzeichnet und anhand dieser Aufzeichnungen die Trefferlisten von Suchanfragen optimiert.

Grundlegende Voraussetzung ist dabei, dass ein Nutzer gegenüber dem Anbieter entsprechend identifiziert werden kann, sodass sonst zustandslose Suchanfragen protokolliert und zum Aufbau eines Profils herangezogen werden können. JEH und WIDDOM entwickeln zu diesem Zweck einen Algorithmus basierend auf dem *PageRank*, welcher einen *personalisierten PageRank-Vektor* ermittelt [7]. Dieser Vektor repräsentiert Präferenzen des Nutzers und gibt Aufschluss über die persönliche Relevanz eines Sets von Webseiten. Mithilfe dieses Vektors werden Trefferlisten gemäß den Nutzerpräferenzen angepasst. Das System passt sich dabei ausgehend von einer initialen Verteilung fortwährend dem Suchverhalten des Nutzers an. Statt die Treffer *explizit* vom Nutzer bezüglich ihrer Relevanz zu bewerten, wird die Gewichtung aus dessen Verhalten abgeleitet – man spricht bei dieser Methode von *impliziter Nutzermodellierung* [11].

Ein zentraler Vorteil dieses Vorgehens ist, dass dem Nutzer kein zusätzlicher Aufwand entsteht. Das Logging der Suchanfragen kann dabei vollautomatisch im Hintergrund ablaufen. Problema-

tisch am Aufbau dieses *Top-Level-Profils* ist, dass sämtliche Suchanfragen eines Nutzers äquivalent behandelt werden. So können beispielsweise unterschiedliche Suchkontexte (*Was tut der Nutzer gerade? Welchem Zweck dient das Informationsbedürfnis?*) nicht berücksichtigt werden [→j].

Personalisierung durch Logging muss nicht zwingend serverseitig durchgeführt werden. Insbesondere im akademischen Umfeld gibt es eine Vielzahl prototypischer Implementierungen, welche Logs auf dem Clientrechner ausführen und Trefferlisten von Suchmaschinen sortieren [9, 12, 2]. Kommerziell durchgesetzt hat sich diese Art der personalisierten Suche bisher jedoch nicht.

3.2 Personalisierung durch Nutzerkollaboration

Die zweite Methodengruppe der Personalisierung fasst unter dem Begriff der *Nutzerkollaboration* eine Reihe verschiedener Konzepte zusammen, die als Gemeinsamkeit die Nutzerpartizipation und Interaktion in den Vordergrund stellen [6]. Unter dem Begriff *social search engine* findet sich unter [→k] eine Klassifikation aus Marktperspektive. Unterschieden werden hier unter anderem *shared bookmarks*, *collaborative directories*, und *personalized verticals*. *Shared bookmarks* sind Dienste, die Nutzern die Möglichkeit bieten, eigene Favoriten im Netz an zentraler Stelle zu speichern und anderen Nutzern zur Verfügung zu stellen. Der Rang einer Webseite innerhalb der Trefferliste hängt dabei von der Zahl der Bookmarks auf diese Seite ab. Das Prinzip ist also dem PageRank sehr ähnlich, legt aber den Fokus der Gewichtung auf die Contentnutzer statt auf die Provider. Beispiele für solche Dienste sind Yahoo! MyWeb oder deli.cio.us. Unter *collaborative directories* werden einfache nutzergenerierte Verzeichnisse verstanden, welche über Suchbegriffe zugänglich sind. Ein Beispiel für einen solchen Dienst ist PreFound. *Personalized verticals* bezeichnen solche Dienste, bei denen sich Nutzer eigene, spezialisierte Suchmaschinen erstellen können. Beispiele für solche Angebote sind Rollyo oder Eurekster. Die Dienste setzen dabei auf etablierte Suchmaschinen auf (im Falle von Rollyo und Eurekster auf die Yahoo! Search Engine).

Eine tatsächliche *Personalisierung der Suche* findet streng genommen nur bei den personalized verticals statt. Im Gegensatz zur Personalisierung durch Logging wird das persönliche Profil jedoch vom Nutzer aktiv gestaltet. Bei den anderen der o. g. Technologien findet eine Personalisierung der Suche nur indirekt über die Assoziation des Nutzers zu anderen statt. Da Nutzer jedoch auch dort eigene Informationen bereitstellen können, wurden diese Dienste in die Betrachtung einbezogen.

4 Gefahren

4.1 Datensicherheit und Missbrauch

Personalisierte Suche in ihren unterschiedlichen Ausprägungen beinhaltet die Preisgabe von Informationen durch den Nutzer, um mithilfe dieser Informationen Suchanfragen semantisch anzureichern. Problematisch jedoch ist, dass insbesondere bei den protokollbasierten Technologien für den Nutzer in der Regel nicht sichtbar ist, welche seiner Daten persistiert werden, in

welcher Form diese auswertbar sind und wie sicher diese Daten vor nicht-autorisiertem Zugriff sind.

SHEN ET AL. unterscheiden vier Level des Schutzes privater Informationen (in erster Linie in Bezug auf Personalisierung durch Logging) [13]:

- *Level I* erfordert eine Pseudoidentität für die Anfragen an die Suchmaschine. Alle Anfragen eines Nutzers werden dabei der gleichen Pseudoidentität zugeordnet und erlauben so eine Personalisierung der Websuche. Die Identifikation des Nutzers kann lediglich indirekt über die Suchanfragen erfolgen.
- *Level II* erfordert die Zusammenfassung von Suchaktivitäten unter einer Gruppenidentität. Mit dieser Sicherheitsstufe geht die Individualität der Suche verloren, jedoch wird eine Gruppe von Nutzern mit ähnlichen Informationsbedarfen in ähnlicher Weise wie ein Einzelnutzer von der Personalisierung profitieren können.
- *Level III* erfordert, dass dem Server keinerlei (direkte oder implizite) Informationen zur Identität des Nutzers übermittelt werden – d. h. auch keine Protokollierung des Suchverhaltens stattfindet. Eine Implikation dieser Forderung ist, dass die Personalisierung der Suchergebnisse nicht mehr serverseitig erfolgen kann. Als Alternative muss am Client spezialisierte Software eingesetzt werden, welche die Ergebnisse der anonym abgesetzten Suchanfragen anhand eines lokal gespeicherten Profils auswertet und sortiert (vgl. Abschnitt 3.1). Die Anonymisierung der einzelnen Suchanfragen gegenüber dem Server lässt sich bspw. über die Nutzung eines TOR-Netzwerkes bewerkstelligen.
- *Level IV* als höchste Stufe der Sicherheit privater Daten erfordert, dass keine Informationen über die Suchanfrage beim Server vorliegen. Dies könnte nur dann gewährleistet werden, wenn der Server des Anbieters nachgewiesenermaßen keine Daten über Suchanfragen persistiert.

Insbesondere die Logging-Mechanismen stellen potenzielle Gefahren für die Privatsphäre dar, da der individuelle Nutzer nur eingeschränkte Möglichkeiten hat, die gespeicherten Informationen zu beeinflussen. Eine weitere Gefahrenquelle, welche über die persönliche Suche hinaus geht, stellt die Konzentration von persönlichen Informationen an einer zentralen Stelle dar [4, 3]. So ist es ggf. möglich, über eine Google-Nutzer-ID auf die Suchhistorie, den privaten E-Mail-Verkehr (GoogleMail) und private Fotos (Google Picasa) zuzugreifen, sofern der Betroffene alle diese Dienste nutzt. Ein nicht autorisierter Zugriff (z. B. über „Erraten“ eines schwachen Passworts oder die Nutzung einer technischen Schwachstelle) hat dabei ein immenses Schadenspotenzial, Gleiches gilt für einen potenziellen Missbrauch durch den Suchmaschinenanbieter selbst.

4.2 Inhaltefilterung

Die Filterung von Inhalten auf Grundlage vorhergehender Suchanfragen oder unter Nutzung eigener, fokussierter Suchmaschinen birgt die grundsätzliche Gefahr der Einseitigkeit. Insbesondere für den Nutzer nicht direkt sichtbare Mechanismen wie die impliziter Nutzermodellierung (vgl. Abschnitt 3.1) könnten dazu führen, dass bevorzugt bekannte Informationsquellen zum Retrieval herangezogen werden. Da der Nutzer den Grad der Beeinflussung der Sucher-

gebnisse bei dieser Form der Personalisierung nicht direkt wahrnimmt, könnte hier eine falsche Einschätzung der Neutralität der Ergebnisse entstehen.

5 Marktübersicht

5.1 Suchmaschinen mit Logging

Das *Google Webprotokoll* (auch *PSearch* bzw. *Google Web History* ist ein von Google gestarteter Suchdienst [→e], welcher sämtliche Suchaktivitäten auf dem Google-Suchportal aufzeichnet und diese bei Suchanfragen berücksichtigt. Der Dienst folgt somit dem Prinzip der *impliziten Nutzermodellierung* (vgl. Abschnitt 3.1). Die Technologie wurde 2003 mit dem kalifornischen Start-up-Unternehmen *Kaltix* eingekauft.

Ziel des Webprotokolls ist die Verbesserung der Trefferqualität anhand nutzerbezogener Kontextinformationen. Voraussetzung ist ein G-Mail-Account, mit welchem der Nutzer angemeldet sein muss. Ist das optionale Protokoll aktiviert, wird jede Suchanfrage sowie jeder angeklickte Link in der Trefferliste als Bookmark protokolliert. Der Nutzer hat jederzeit die Möglichkeit, dieses Log einzusehen (vgl. Abbildung 1) und auch einzelne Einträge daraus zu löschen. Mit dem Browser-Plugin „Google Toolbar“ können Webseiten auch direkt in das Protokoll geschrieben werden.

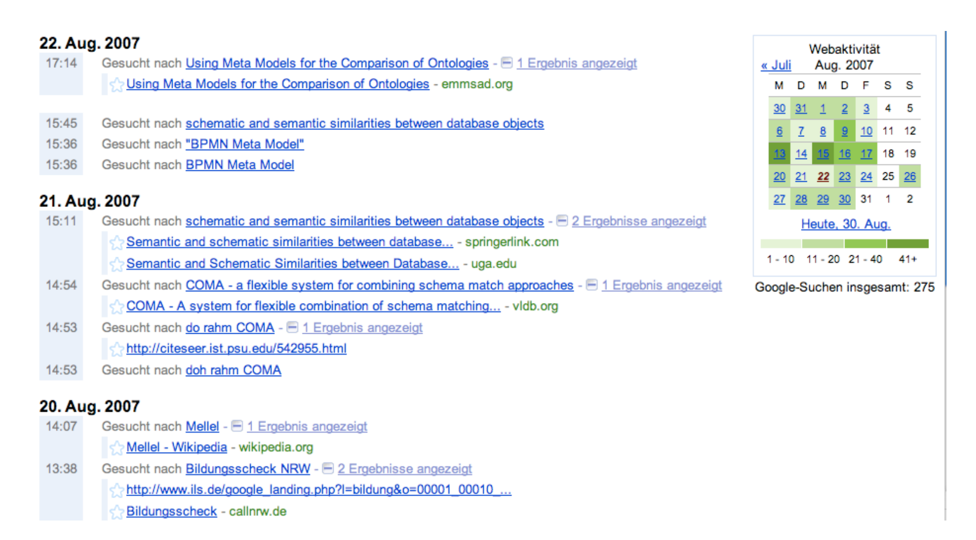


Abbildung 1: Das Google-Webprotokoll

Für den Nutzer wird jedoch nicht transparent, in welcher Art das Webprotokoll die Suche beeinflusst. Eine Umfrage unter den Lesern von *readwriteweb.com* ergab, dass die überwiegende Mehrheit den Unterschied als gering oder nicht spürbar einschätzt, von einigen Nutzern wurde eine Verschlechterung der Suchergebnisse beobachtet [→f]. Eigene Experimente im Rahmen der Arbeit brachten ebenfalls kaum Unterschiede zwischen personalisierten und anonym abgesetzten bzw. auch zwischen dediziert unterschiedlich trainierten Accounts zutage.

5.2 Suchmaschinen mit Nutzerkollaboration

Yahoo! Search MyWeb

Hinter dem Begriff *Yahoo! MyWeb* [→h] verbirgt sich ein Dienst vom Typ *shared bookmarks* (bzw. *social bookmarking*). Nutzer von Yahoo! MyWeb können dabei ihre Favoriten auf dem Portal des Anbieters speichern, mit Tags versehen und mit anderen Nutzern teilen. Streng genommen findet also keine Personalisierung der Suche statt. Die Technologie der Plattform wurde vom Social-Bookmarking-Anbieter deli.cio.us [→g] übernommen, den Yahoo! im Dezember 2005 gekauft hat. Statt MyWeb und deli.cio.us zu integrieren, lässt Yahoo! die Dienste nebeneinander existieren – wohl hauptsächlich, um die Nutzerbasis von deli.cio.us zu erhalten [→i].

Rollyo

Rollyo (Beta) [→l] ist eine Suchmaschine, die auf einem Social-Search-Prinzip (bzw. *personalized vertical*) basiert und seit September 2005 online ist. Die Nutzer des Angebots können so genannte *Search Rolls* anlegen. Das sind gespeicherte Suchanfragen, welche eine Liste von „trusted websites" enthalten, die der Nutzer festlegt. Diese „nutzerdefinierte Suchmaschine" nimmt anschließend Suchanfragen entgegen, die nur auf den vorher festgelegten Webseiten ausgewertet werden. Dazu nutzt Rollyo die Engine von Yahoo! Search. Die Search Rolls können veröffentlicht und mit Tags annotiert werden, anhand derer andere Nutzer diese auffinden können.

Eurekster swicki

Eurekster swicki (beta) [→m] ist wie Rollyo eine Suchmaschine nach dem Prinzip *personalized vertical*, die seit November 2005 online ist. Sie setzt ebenfalls auf der Engine von Yahoo! Search auf. Ein *Swicki* ist ähnlich einer Rollyo Search Roll eine benutzerdefinierte Suchmaske, die von den Nutzern der Webseite angelegt und veröffentlicht wird. Dabei legt der Nutzer eine Reihe von Suchwörtern für das Swicki an. Diese werden dann zu jedem Suchbegriff, der in die Swicki-Maske eingegeben wird, hinzugefügt und der Suchstring dann an die Yahoo-API weitergereicht. Ein Swicki kann darüber hinaus wie die Search Rolls eine Liste von präferierten Seiten enthalten, deren Treffer bevorzugt behandelt werden, sowie spezifische Seiten blocken. Die Swicki-Technologie bietet darüber hinaus eine adaptive Filterung der Suchergebnisse anhand des Suchverhaltens der Nutzergemeinde. Als besonderes Feature hat jedes Swicki eine so genannte *Tag Cloud*, welche häufig benutzte Suchbegriffe des Swickis visualisiert. Die Zielgruppe für Eurekster sind hauptsächlich Webmaster und Blogautoren, da sich Swickis insbesondere zur Integration in die persönliche Webseite eignen. Mithilfe der Seitenpriorität kann so den Nutzern der eigenen Webseite eine Suche angeboten werden.

PreFound

Bei *PreFound* [→n] handelt es sich streng genommen nicht um eine Suchmaschine, sondern um einen nutzerbetriebenen Verzeichnisdienst, welcher Anfang 2006 online ging. Die Nutzer von PreFound erzeugen dabei Verzeichnisse (Groups), benennen diese und versehen sie mit einer Reihe von Tags. Anhand dieser Tags können die Verzeichnisse dann über die seiteninterne Suche von anderen Nutzern gefunden und genutzt werden. Die Verzeichnisse können durch die Nutzer bewertet werden – die Bewertung fließt in das Ranking der Trefferliste ein.

Anders als die bisherigen vorgestellten Anbieter personalisierter Suche, setzt PreFound nicht auf einen WebCrawler wie Google oder Yahoo! auf, sondern beruht auf einer vollständig auf Nutzereingaben basierenden Indizierung. Zum Zeitpunkt dieses Beitrags war die Zahl der Einträge für allgemeine Begriffe auf PreFound eher gering und die Abdeckung spezieller Suchbegriffe sehr lückenhaft, was sicher auch auf die relativ kurze bisherige Existenz des Dienstes zurückzuführen ist.

Aftervote

Aftervote [→o] ist eine sehr komfortable Metasuchmaschine, die die Suchmaschinen Google, Yahoo! und MSN Live Search durchsucht. Das Angebot ist seit April 2007 online. Das Ranking der Suchergebnisse hängt zum einen vom Ranking auf den Quellsuchmaschinen ab, andererseits haben Nutzer per „vote up“ die Möglichkeit, bestimmten Links eine höhere Relevanz zuzuordnen.[2] Die Votes werden entsprechend in die Rankings zukünftiger Anfragen einbezogen. Welche Rankings mit welchem Gewicht in die Generierung einer Trefferliste eingehen, lässt sich vom Nutzer individuell einstellen. Darüber hinaus lassen sich Blacklists erstellen, welche ausgewählte Domains bei zukünftigen Anfragen blockieren mit Whitelists lassen sich Domains für die Trefferanzeige bevorzugen. Eine thematische Personalisierung der Suche ist jedoch nicht möglich.

Wird ein Link in der Trefferliste angeklickt, so wird nicht direkt zum jeweiligen Webserver gewechselt, sondern die gefundene Seite wird in einer Art Frame angezeigt, welcher am unteren Rand zahlreiche Funktionalitäten zur Verfügung stellt. Darunter sind neben der Möglichkeit, die Seite zu bewerten, auch zahlreiche PlugIns für Social-Bookmarking-Dienste wie *deli.cio.us* oder *digg*.

6 Fazit und Ausblick

Der Suchmaschinenmarkt im Rahmen der Web 2.0-Bewegung zeichnet sich insbesondere durch zahlreiche neue Angebote aus, die verstärkt auf die Partizipation einer Nutzergemeinschaft setzen. Die Personalisierung der Nutzung dieser Dienste ist dabei meist unumgängliche Voraussetzung – zumindest eine Identifikation des Nutzers durch eine (Pseudo-)Identität ist gängige Praxis.

Welche dieser Anbieter sich langfristig durchsetzen werden können bzw., ob das Prinzip „Social Search“ überhaupt eine Zukunft hat, ist aufgrund der relativ jungen Konzepte noch nicht abzuschätzen. Das gilt für die kleinen Start-ups ebenso wie für experimentelle Dienste der etablierten Anbieter. So hat bspw. Google Ende 2006 seinen Dienst *Answers* nach vier Jahren eingestellt.

Insbesondere die Social-Search-Dienste sind stark von ihrer Nutzerbasis abhängig, sollen sie einen echten Mehrwert gegenüber traditionellen Suchdiensten haben – hier konkurrieren die vielfältigen Angebote miteinander. Abschließend lässt sich beobachten, dass sämtliche der untersuchten Dienste entweder direkt auf etablierte Suchmaschinentechnologie aufsetzen oder

---

2 Der Dienst url.com [→p] funktioniert nach dem gleichen Prinzip.

auf eine Nutzerbasis vertrauen, die ihre Inhalte selbstständig auf Grundlage dieser Kanäle erstellen. Langfristig durchsetzen wird sich also wahrscheinlich das Angebot, das traditionelle Suche und Personalisierungsaspekte aus Nutzersicht am Besten zu integrieren versteht.

Linkverzeichnis

[→a] http://www.google.com; http://www.google.de
[→b] http://www.yahoo.com; http://www.yahoo.de
[→c] http://www.live.com; http://www.live.de
[→d] http://www.ask.com; http://www.ask.de
[→e] http://www.google.com/history
[→f] http://www.readwriteweb.com/archives/how_effective_is_google_personalized_search.php
[→g] http://deli.cio.us
[→h] http://myweb.yahoo.com
[→i] http://www.readwriteweb.com/archives/top_10_yahoo_properties.php
[→j] http://www.readwriteweb.com/archives/personalized_search_primer.php
[→k] http://searchenginewatch.com/showPage.html? page=3623173
[→l] http://www.rollyo.com
[→m] http://www.eurekster.com
[→n] http://www.prefound.com
[→o] http://www.aftervote.com
[→p] http://www.url.com

Literaturverzeichnis

[1] Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., Su, Z. Optimizing web search using social annotations. In WWW ’07: Proceedings of the 16th international conference on World Wide Web (New York, NY, USA, 2007), ACM Press, pp. 501-510.

[2] Cassel, L. N., Wolz, U. W. U., Beck, R. E. Server independent personalization for effective searching in the world wide web. In AH (2002), P. D. Bra, P. Brusilovsky, R. Conejo, Eds., vol. 2347 of Lecture Notes in Computer Science, Springer, pp. 488-491.

[3] Conti, G. Googling considered harmful. In NSPW ’06: Proceedings of the 2006 workshop on New security paradigms (New York, NY, USA, 2007), ACM Press, pp. 67-76.

[4] Conti, G., Sobiesk, E. An honest man has nothing to fear: user perceptions on web-based information disclosure. In SOUPS ’07: Proceedings of the 3rd symposium on Usable privacy and security (New York, NY, USA, 2007), ACM Press, pp. 112-121.

[5] de Kunder, M. The size of the world wide web, daily estimated size of the world wide web. http://www.worldwidewebsize.com, 2007.

[6] Freyne, J., Farzan, R., Brusilovsky, P., Smyth, B., Coyle, M. Collecting community wisdom: integrating social search & social navigation. In IUI ’07: Proceedings of the 12th international conference on Intelligent user interfaces (New York, NY, USA, 2007), ACM Press, pp. 52-61.

[7] Jeh, G., Widom, J. Scaling personalized web search. In WWW ’03: Proceedings of the 12th international conference on World Wide Web (New York, NY, USA, 2003), ACM Press, pp. 271-279.

[8] Kobayashi, M., Takeda, K. Information retrieval on the web. ACM Comput. Surv. 32, 2 (2000), 144-173.

[9] Lv, Y., Sun, L., Zhang, J., Nie, J.-Y., Chen, W., Zhang, W. An iterative implicit feedback approach to personalized search. In ACL ’06: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL (Morristown, NJ, USA, 2006), Association for Computational Linguistics, pp. 585-592.

[10] Page, L., Brin, S., Motwani, R., Winograd, T. The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford Digital Library Technologies Project, 1998.

[11] Shen, X., Tan, B., Zhai, C. Implicit user modeling for personalized search. In CIKM ’05: Proceedings of the 14th ACM international conference on Information and knowledge management (New York, NY, USA, 2005), ACM Press, pp. 824-831.

[12] Shen, X., Tan, B., Zhai, C. Ucair: a personalized search toolbar. In SIGIR ’05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (New York, NY, USA, 2005), ACM Press, pp. 681-681.

[13] Shen, X., Tan, B., Zhai, C. Privacy protection in personalized search. SIGIR Forum 41, 1 (2007), 4-17.

[14] Walker, J. Links and power: the political economy of linking on the web. In HYPERTEXT ’02: Proceedings of the thirteenth ACM conference on Hypertext and hypermedia (New York, NY, USA, 2002), ACM Press, pp. 72-73.

[15] White, R. W., Drucker, S. M. Investigating behavioral variability in web search. In WWW ’07: Proceedings of the 16th international conference on World Wide Web (New York, NY, USA, 2007), ACM Press, pp. 21-30.

[16] Xing, B., Lin, Z. The impact of search engine optimization on online advertising market. In ICEC ’06: Proceedings of the 8th international conference on Electronic commerce (New York, NY, USA, 2006), ACM Press, pp. 519-529.

Blogging vs. Knowledge Management – Wie Blogs zu gutem Wissensmanagement in Organisationen beitragen können

Daniel Beverungen

1 Einleitung

Wissen stellt für Unternehmen zunehmend eine entscheidende strategische Ressource dar. Wissensmanagement in Organisationen soll in diesem Zuge zu einem systematischen Umgang mit der Ressource Wissen führen.

Auf der anderen Seite wird unter dem Modebegriff Web 2.0 eine Vielfalt neuer Dienste subsumiert, die zu einer partizipativen Nutzung des Internets anregen. In diesem Zuge sind beispielsweise Weblogs (Blogs) Onlinetagebücher, die kooperativ erweitert werden können, zu nennen. Corporate Blogs als Spezialfall bezeichnen Blogs, die durch Unternehmen – entweder zur internen Kommunikation oder auch zur Kommunikation mit externen Akteuren wie Kunden, Lieferanten etc. – betrieben werden.

Im Schnittbereich der Konzepte stellt sich somit die Frage, ob und inwiefern Corporate Blogs das Wissensmanagement in Organisationen fördern und ergänzen können und welche Maßnahmen vonseiten des Managements und der Belegschaft zu einer sinnvollen Nutzung von Blogs ergriffen werden können.

Im Rahmen dieser Ausarbeitung werden somit beide Konzepte kurz charakterisiert (Kapitel 2). Nachfolgend wird anhand der Phasen der Wissensschaffung nach NONAKA und TAKEUCHI untersucht, wie Corporate Blogging das Wissensmanagement in Organisationen erweitern und verbessern können. Auch erfolgt eine Betrachtung von Blogs vor dem Hintergrund der Wissensmanagementphasen nach PROBST, RAUB und ROMHARDT. Vermutete Managementimplikationen und Erfolgsfaktoren werden ebenfalls vorgestellt (Kapitel 3). Eine Zusammenfassung der Ergebnisse erfolgt in Kapitel 4.

2 Kurzvorstellung der Konzepte

2.1 Knowledge Management

Wissensmanagement (engl. *Knowledge Management*) ist ein Managementkonzept zur systematischen Erstellung, Erhaltung und Weiterentwicklung des Umgangs einer Organisation mit ihrem individuellen und gemeinsamen Wissen, um ihre strategischen Ziele zu erreichen.[1]

---

1 Vgl. Bennet, Bennet (2003), S. 440.

Der Begriff Wissen wird dabei häufig von den Begriffen Information und Daten abgegrenzt.[2] Demnach sind Daten nach syntaktischen Regeln verknüpfte Symbole[3] ohne semantische Bedeutung (z. B. 7438). Daten, die von einem Individuum wahrgenommen werden und in diesem Zuge einen „Sinn“ ergeben, bezeichnet man als Information (z. B. Dax wurde am 21.08.2007 mit 7438 Punkten notiert). Wissen besteht aus miteinander vernetzten und bewerteten Informationen. In der Literatur wird häufig die Ansicht vertreten, dass Wissen nur im Individuum existieren kann; daher ist Wissen immer an die es besitzende Person gebunden. Durch die „hierarchische“ Differenzierung von Daten, Informationen und Wissen umfasst Wissensmanagement daher eine systematische Behandlung von Wissen, Informationen und Daten.

Unterschieden wird zwischen explizitem und implizitem Wissen (engl. *tacit knowledge*). Implizites Wissen liegt unausgesprochen vor (z. B. Form eines Adlerflügels), explizites Wissen liegt kodiert vor. Eine Möglichkeit hierzu ist die Explikation von Wissen mittels einer Auswahl von Informationsartefakten (z. B. gesprochene Worte, Dateien oder Schriftstücke). Jedoch ist implizites Wissen nicht grundsätzlich explizierbar, sondern kann häufig nur teilweise expliziert werden (z. B. die genaue Vorgehensweise zu einem ansprechendem Vortrag von Beethovens 5. Symphonie).

Vor diesem Hintergrund beschreibt Abbildung 1 den Wissensaustausch zwischen zwei Individuen mittels entsprechender Informationsartefakte.[4] Die „wissende“ Person hat eine bestimmte Vorstellung eines Sachverhaltes (sog. mentales Modell). Um der anderen Person dieses Wissen zukommen zu lassen, wird es auf Informationsebene heruntergebrochen. Die Informationen werden in bestimmte Artefakte kodiert und an die andere Person übermittelt, die in einem Dekodierungsprozess die Information extrahiert und in sein eigenes Wissen einbettet. Moderierend wirkt eine (Unternehmens-)Kultur, die gemeinsame Werte und Begriffe zur Verfügung stellt, wodurch insbesondere der Dekodierungsprozess erleichtert wird.

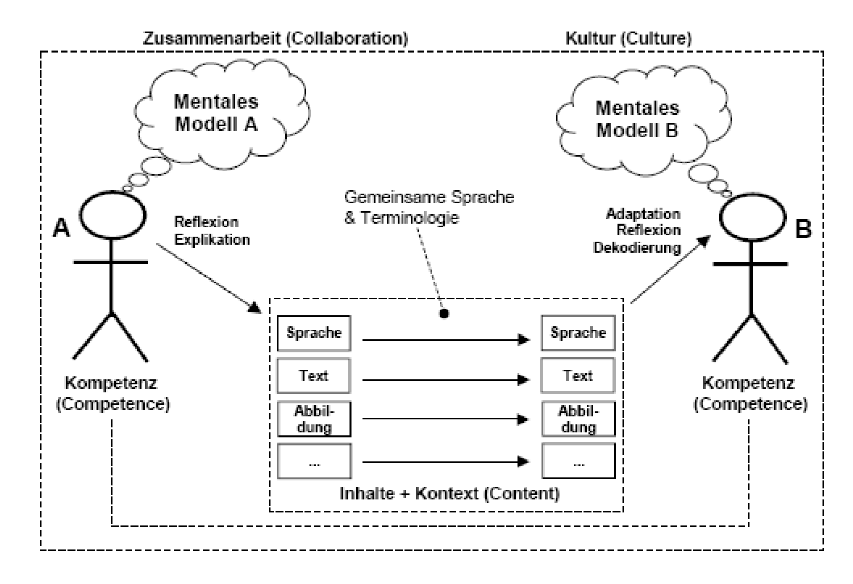


Abbildung 1: Prozess des Wissensaustausches mittels Informationsartefakten

2 Andere Systematiken verstehen jedoch auch Daten und Informationen als Wissen, so z. B. das von van Lohuizen vorgeschlagene „Complete set of knowledge states“, vgl. Holsapple (2003), S. 168, oder schlagen ein Kontinuum aus Daten, Informationen und Wissen vor, z. B. Probst, Raub, Romhardt (2003), S. 17.

3 Vgl. DIN (2006).

4 Vgl. hierzu und zu Abbildung 1 in Riempp (2004), S. 64-66.

Innerhalb von Organisationen liegt Wissen folglich implizit in den Köpfen der Mitarbeiter sowie expliziert in Informationsartefakten vor. Die Gesamtheit dieses Wissens stellt das sog. organisationale Gedächtnis (engl. *Organizational Memory*) dar.

Wie eingangs erwähnt, ist Wissensmanagement ein Managementkonzept zum systematischen Umgang mit dem Wissen der Organisation. Unterscheiden lassen sich dabei folgende Aktivitäten, die sich nach PROBST, RAUB und ROMHARDT[5] zu einem Kreislauf ergänzen (vgl. Abbildung 2). Das Wissensmanagement folgt dabei einer Wissensmanagementstrategie, die im Einklang mit der Business-Strategie des Unternehmens aufzustellen ist. Über die Identifikation erforderlichen Wissens und den Erwerb (z. B. von anderen Unternehmen, Stakeholdern oder externen Wissensträgern) hinaus, wird Wissen weiterentwickelt, in der Organisation selektiv verteilt und kommuniziert und schließlich zur Erfüllung bestimmter Ziele genutzt. Zur Vermeidung eines „Gedächtnisverlustes“ aufbewahrt, erfolgt daraufhin eine Bewertung des vorhandenen gegenüber dem erforderlichen Wissen. Ein etwaiges Delta führt zu einer entsprechenden Anpassung der Wissensziele.

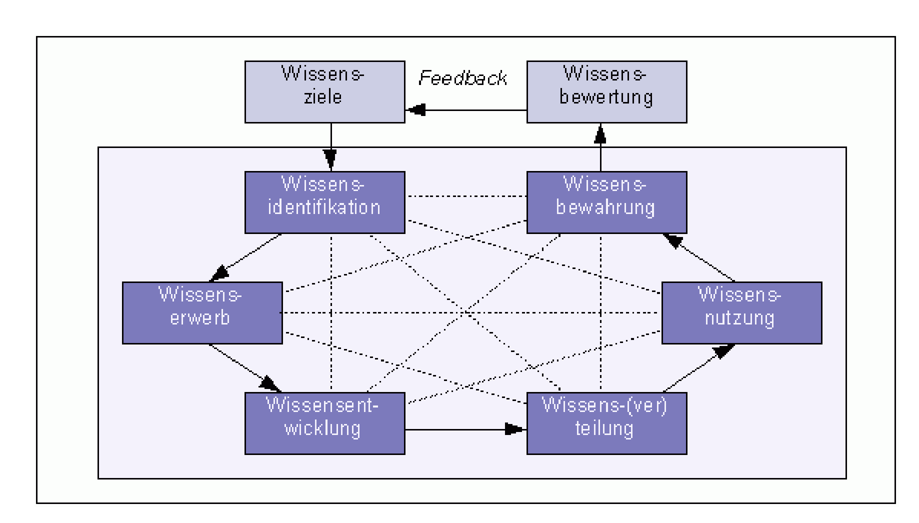


Abbildung 2: **Wissenskreislauf nach PROBST, RAUB und ROMHARDT[6,7]**

Eng verzahnt mit dem Wissensmanagement sind andere Forschungsgebiete (vgl. Abbildung 3). Das Konzept der lernenden Organisation (engl. *Learning Organization*) stellt dabei eine andere Sicht auf das Wissen einer Organisation dar: Organisationen sind zum Lernen auf das Vorhandensein einer Wissensbasis angewiesen, die im Zuge des Lernprozesses angepasst (Verändern), verkleinert (Vergessen) oder erweitert (Dazulernen) wird. Lernen bedeutet folglich eine Veränderung im organisationalen Gedächtnis. Ein nicht funktionierender Wissensaustausch in Organisationen wird daher nicht selten mit einer „Lernschwäche“ der Organisation gleichgesetzt.[8]

---

5 Vgl. Probst, Raub, Romhardt (2003).

6 Vgl. Probst, Raub, Romhardt (2003).

7 Das Bild stamm von: http://www.wissensmanagement.net/online/archiv/2004/08_2004/wm-bausteine.shtml.

8 Zum organisationalen Lernen vgl. z. B. Argyris (1993) oder Argyris, Schön (1978).



Abbildung 3: Wissensmanagement und verwandte Disziplinen[9]

Wissenstransformation: Das Konzept der Wissenstransformationen zwischen explizitem und implizitem Wissen – ursprünglich entwickelt durch NONAKA und TAKEUCHI[10] – hat in der Literatur große Beachtung gefunden. Im Modell werden Prozesse veranschaulicht, die bei der Schaffung und Verbreitung von Wissen in Unternehmen anzutreffen sind. Im Einzelnen sind dies:

Sozialisation[11]: Dies ist der Austausch impliziten Wissens, der durch die gemeinsam gemachten Erfahrungen und Beobachtungen der Handlungen anderer stattfindet. Dadurch entwickeln sich gemeinsame Fertigkeiten und Fähigkeiten sowie gemeinsame mentale Modelle.[12]

Externalisierung: Implizites Wissen wird in Form von Metaphern, Analogien, Konzepten und Begriffen, Hypothesen und Modellen expliziert. Dabei kann eine Reflexion des Wissens erfolgen, die einen Verbesserungsprozess fördert.[13]

Kombination: Explizites Wissen aus unterschiedlichen Bereichen kann kombiniert, sortiert, klassifiziert und zu anderem Wissen hinzugefügt werden.

Internalisierung: Dies ist die Aufnahme von explizit vorhandenen Wissensbestandteilen durch Individuen und gewissermaßen der Rückfluss von Wissen aus der organisationalen Wissensbasis zur individuellen Wissensbasis durch das sog. *Learning by Doing*: Durch die Anwendung wird Wissen Teil der mentalen Modelle sowie der Fertigkeiten und Fähigkeiten[14] des Individuums.

Diese Prozesse bilden zusammen einen Kreislauf, der unter Anwendung des Wissensmanagements zu organisationalem Lernen führt (vgl. Abbildung 4).

---

9 Quelle: Forschungszentrum für Informatik (2007): http://www.fzi.de/ipe/1904.html.

10 Vgl. Nonaka, Takeuchi (1995), S. 62-73.

11 oder auch: Sozialisierung.

12 Vgl. Krcmar (2005), S. 482.

13 Vgl. Krcmar (2005), S. 483.

14 Vgl. Krcmar (2005), S. 484.

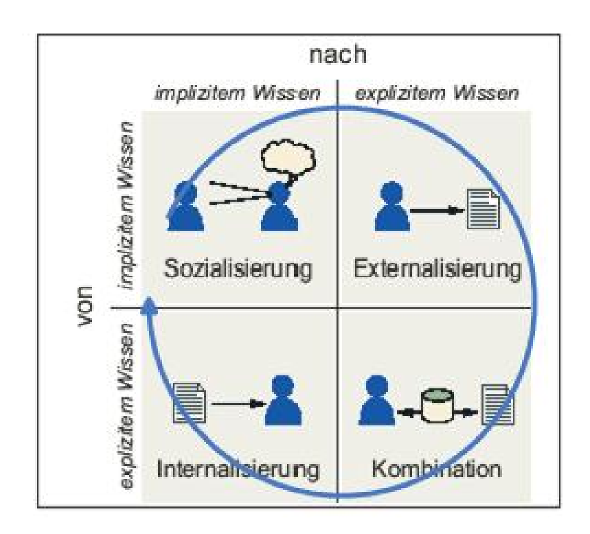


Abbildung 4: Prozesse der Wissensschaffung und -verbreitung in Organisationen

2.2 Blogging im Kontext von Web 2.0

Der Begriff Web 2.0 ist ein ursprünglich von TIM O'REILLY propagierter Modebegriff,[15] der den Wandel des Internets von der individuellen Nutzung von Informationen zum eher partizipatorischen Umgang beschreibt.[16] In diesem Zuge ist Web 2.0 sowohl durch die Bereitstellung entsprechender Dienste als technologischer Enabler als auch sozialer Wegbereiter einer „Mitmach-Kultur" im Internet zu verstehen. Blogs stellen (seit der Bereitstellung der entsprechenden Software 1999[17]) in diesem Zuge eine der Erscheinungsformen des Web 2.0 dar, da sowohl die technologische Umsetzbarkeit durch vorgefertigte Lösungen bereitgestellt wird als auch explizit das gegenseitige Weiterentwickeln und Überarbeiten der Inhalte im Sinne einer kollektiven Intelligenz der Nutzer[18] ein Teil des Konzeptes ist. Das Thema Web 2.0 erfährt aktuell ein hohes Maß an Aufmerksamkeit und wird momentan an der Spitze des Gartner Hype-Cycles gesehen.[19]

„Blog" ist ein Kunstwort, das sich aus den Bestandteilen „Web" und „Logbuch" zusammensetzt und eine Webseite im Internet bezeichnet, in dem fortlaufend neue Einträge[20] – meist in umgekehrter chronologischer Reihenfolge – vorgenommen werden. Neben den aktuellen Einträgen sind weitere Einträge häufig in einem beigeordneten Archiv abgelegt. Als „Blogger" bezeichnet man die Autoren der Beiträge eines Blogs. Die Beiträge können dabei von einer Person allein

---

15 Vgl. HMD 2007, S. 116.

16 Vgl. Cyganski (2007).

17 Vgl. Clyde (2004), S. 390-391.

18 Vgl. Alby (2007), S. 15. Weitere erfolgreiche Strategien des Web 2.0 sind (wie hier genannt): Nutzung des Webs als Plattform, Zugang zu Daten, die schwer oder teuer zusammenzustellen sind und wertvoller werden – je häufiger sie genutzt werden – nutzerorientierte Softwareentwicklung, leichtgewichtige Modelle, Software, die die Grenzen einzelner Geräte überschreitet und das Einbeziehen sog. Long-Tail-Effekte durch Selfservice.

19 So dargestellt in der Zeitschrift Computerwoche, Jg. 38, Ausgabe 32-33 (2007), S. 4.

20 Vgl. HMD 2007, S. 116.

oder kooperativ von mehreren Personen (sog. *Group Blogs*) erstellt oder kommentiert[21] werden.[22]

Blogs können unterschiedlichste Themengebiete behandeln,[23] z. B. politisch motivierte Blogs,[24] satirische Blogs über Peinliches aus dem Privatleben von Prominenten[25] oder Informationen über aktuelle Trends und technologische Unterstützung im Internet[26] (vgl. Abbildung 5). ALBY[27] nennt als wichtigste Ausprägungsformen von Blogs:

- *Watchblogs*: Kritische Beobachtung von Medien und Firmen
- *Liblogs*: Blogs über Literatur
- *Coporate* Blogs: Blogs von Firmen
- *Blawgs*: Blogs mit juristischen Themen
- *Fotoblogs*: Blogs, in denen hauptsächlich Fotos veröffentlicht werden



Abbildung 5: MyBlogLog – Beispiel eines Blogs[28]

Laut einer Einschätzung des Spiegels (2007) nimmt die Verbreitung der Blogs im Internet auf hohem Niveau weiter zu. In den USA wird die Anzahl der Blogs derzeit auf ca. 10 Mio. beziffert, in Deutschland beträgt diese Zahl momentan ca. 100.000. Im Einzelfall ist jedoch die Resonanz

---

21 Vgl. Alby (2007), S. 22.

22 Vgl. Clyde (2004), S. 391.

23 Zu näheren Informationen über die Beispiele vgl. Spiegel (2007).

24 www.hoder.com, www.sobhaneh.com (arabisches Original): Oppositionelle Kritik der iranischen Staatsführung mit bis zu 40.000 Lesern täglich, vgl. Spiegel (2007), S. 30.

25 http://gofugyourself.typepad.com/.

26 http://www.readwriteweb.com/, http://www.mybloglog.com/buzz/community/adaptiveblue/, www.myBlog.de.

27 Vgl. Alby (2007), S. 21.

28 http://www.mybloglog.com/buzz/community/adaptiveblue/.

der entsprechenden Beiträge bei der Leserschaft schwer einzuschätzen. Eine besondere Rolle bei der Qualitätsbeurteilung kommt daher der Vertrauenswürdigkeit des jeweiligen Autors zu.[29]

Bedingt durch die Unüberschaubarkeit der **Blogosphäre[30]** bleibt in Untersuchungen häufig unklar, wie viele Blogs von anderen Nutzern tatsächlich gelesen werden. Stattdessen werden – ganz im Sinne des Web 2.0-Gedankens – alle möglichen Themen auf unterschiedlichstem Niveau behandelt, auch wenn die Leserschaft auf diesem Themengebiet nur eng umgrenzt ist.

Da Blogs folglich stark durch subjektive Einstellungen und Individualität geprägt sind, dürfen sie somit nicht als objektive und fundierte Informationsquellen angesehen werden.[31] Ermöglicht wird dies u. a. auch dadurch, dass nur der Herausgeber des Blogs selbst zur Anstoßung neuer Themen berechtigt ist (was Blogs von Foren unterscheidet),[32] während andere nur kommentieren können. Um ein größtmögliches Maß an Objektivität herzustellen, wäre die Erfüllung diverser Anforderungen notwendig. Eine Reglementierung liefe jedoch aus Sicht vieler Blogger dem Grundgedanken eines freien und unzensierten Meinungsaustausches im Sinne der bereits angesprochenen Web 2.0-Mentalität zuwider.

Technisch erlauben Blogs auch – über sog. Permalinks – die Vernetzung von Blogs untereinander, sodass sich Themen sehr schnell verbreiten können und permanent auffindbar bleiben. Feeds (z. B. der RSS-Feed) ermöglichen es, interessierte Nutzer automatisch auf dem neuesten Stand zu halten.[33] Dagegen ist eine automatische Suche in Blogs derzeit wenig entwickelt oder führt kaum zu brauchbaren Ergebnissen.[34]

3 Corporate Blogging und Verbindungen zum Knowledge Management

Wie in Kapitel 2 gezeigt, bezeichnet das *Corporate Blogging* die Erstellung von Blogs durch Unternehmen. Nachfolgend sollen Corporate Blogs auf ihre Beziehung zum Wissensmanagement in Unternehmen untersucht werden. Diese Beschränkung ergibt sich daraus, dass das Wissensmanagement als Management in Organisationen bzgl. Blogs in erster Linie mit Corporate Blogs in Beziehung zu setzen ist.

Dabei können Corporate Blogs unterschiedliche Zielsetzungen verfolgen.[35] Zu unterscheiden sind hier externe (zu lesen auch außerhalb der Organisation) und interne Blogs.[36] Da externe Blogs hauptsächlich der Pflege von Beziehungen zu Individuen außerhalb der Organisation dienen (z. B. Supplier/Customer Relationship), erfolgt im Folgenden die Betrachtung von internen

---

29 Vgl. Clyde (2004), S. 391.

30 Blogosphäre = Gesamtheit der Blogs, vgl. Alby (2007), S. 22.

31 Vgl. HMD (2007), S. 116.

32 Vgl. Alby (2007), S. 22.

33 Vgl. Alby (2007), S. 23.

34 Vgl. Therwall, Hasler (2007).

35 Eine Auflistung von Corporate Blogs der Fortune 500 Companies findet sich z. B. unter http://www.eu.socialtext.net/bizblogs/index.cgi?fortune_500_business_blogging_wiki, Abruf: 29.08.2007.

36 Vgl. Jackson, Yates, Orlikowski (2007), S. 1.

Blogs und ihre Beziehung zum Wissensmanagement der Organisation. Die Darstellung orientiert sich dabei an den vorgestellten Phasen der Wissensschaffung von NONAKA und TAKEUCHI.

3.1 Sozialisierung

Der soziale Aspekt wird in verschiedenen Publikationen als der herausragende Nutzen des Blogging in Organisationen herausgestellt.[37]

Da Wissen häufig als implizites, nicht explizierbares Wissen in den Köpfen der Mitarbeiter vorliegt, bietet die Sozialisierung der Mitarbeiter die Voraussetzung zur Verbreitung dieses Wissens in der Organisation (durch gemeinsam gemachte Erfahrungen und gemeinsame mentale Modelle).

Insb. die Sozialisierung der Mitarbeiter untereinander wird durch das Kommunizieren (ggf. auch über Hobbies etc.) über Blogs erleichtert; durch den Grundgedanken der offenen und ungehemmten Kommunikation im Sinne des Web 2.0-Gedankens werden soziale Barrieren des Wissensaustausches abgebaut. Somit kann die Schaffung einer Unternehmenskultur, die zur Wissensverbreitung förderlich ist, erleichtert werden.

Mithin fungiert das Blogging als Möglichkeit für Mitarbeiter, ihr soziales Netzwerk in der Organisation zu erweitern und mit Kollegen in Kontakt zu kommen, die man sonst nicht kennengelernt hätte. So können zusätzliche Wissensquellen erschlossen werden, auf die im Bedarfsfall zugegriffen werden kann.[38]

3.2 Externalisierung

Eine Externalisierung von Wissen erfolgt über Informationsträger (z. B. Sprache oder Dateien). Bezogen auf Blogs wird Wissen externalisiert, wenn entsprechend Einträge oder Kommentare in Blogs geschrieben werden. Neben dem Informationsgehalt selbst bietet dieser Eintrag anderen Personen die Möglichkeit, auf das Wissen der betreffenden Person rückzuschließen (im Sinne einer Wissenslandkarte). Somit kann der Blogger im Bedarfsfall zu Rate gezogen werden, wenn sein Wissen für andere von Nutzen sein kann.

Blogs stellen somit eine technische Plattform zur Externalisierung von Wissen bereit. Sie dienen somit zur Äußerung von Ideen, Themen und Meinungen (Leipzig). Dabei sind sie einfacher zu nutzen als andere technische Hilfsmittel (z. B. Webseiten[39]), sodass auch technisch weniger versierte Mitarbeiter zum Teilen ihres Wissens befähigt werden.

Durch die Charakterisierung von Blogs als zeitlich geordnete Kurzbeiträge eignen sich Blogs besonders zur Verbreitung und Diskussion von sog. „Information Nuggets“[40], nicht aber zur Darstellung von Informationen, die sehr stark mit anderen Informationen verknüpft sein müssen um sinnvoll zu sein oder zur Darstellung sehr komplizierter Sachverhalte.

---

[37] Vgl. z. B. Cayzer (2004), S. 48.
[38] Vgl. Jackson, Yates, Orlikowski (2007).
[39] Vgl. Cayzer (2004), S. 48.
[40] Vgl. Cayzer (2004), S. 48.

Durch die (gewollte) Unvorhersehbarkeit der Inhalte (und damit des Informationsgehaltes) des Blogs ist die Qualität des externalisierten Wissens dabei zwischen den einzelnen Einträgen vermutlich sehr unterschiedlich. Im Sinne einer zielgerichteten Externalisierung wäre hier auf einen möglichst hohen Informationsgehalt der Beiträge abzuzielen.

3.3 Kombination

Das Blogging ist ein selbstorganisierender und kreativer Prozess und keinesfalls strukturiert im Sinne eines ganzheitlichen Managementkonzeptes. Die Stärke zur Wissensverbreitung liegt aber auch genau in dieser kreativen Unvorhersehbarkeit des Ergebnisses.

Ein Blog beinhaltet per Definition Einträge in chronologischer Reihenfolge. Daher erscheint es auf den ersten Blick zur Unterstützung eines fortwährend anzupassenden organisationalen Gedächtnisses ungeeignet, weil ältere Beiträge (auch falls sie im Blog archiviert werden) auf dem Fokus verschwinden und kaum weitere Beachtung erfahren. Auch Suchstrategien erscheinen aufgrund des unstrukturierten Aufbaus und der auftretenden Begriffsvielfalt wenig erfolgversprechend. Vielmehr eigenen sich Blogs zu einer zeitlich begrenzten Kombination von punktuell interessantem Wissen über aktuell relevante Themen.

3.4 Internalisierung

Bedingt durch den kreativen Prozess des Blogging ist das jeweilige Ergebnis der Kombination stark von der sinnvollen Teilnahme der „richtigen" Personen abhängig. Mithin sind Blogs – wie in manchen Untersuchungen dargestellt – nur bedingt eine Informationsquelle (z. B. werden Blogs laut einer Umfrage der Uni Leipzig nur von 9,6 % der Befragten als geeignete Informationsquelle bezeichnet) für die Mitarbeiter.

3.5 Schlussfolgerung und Managementimplikationen

Als Ergebnis der oben durchgeführten Untersuchung der Eigenschaften von Blogs bezüglich der Arten der Wissensschaffung in Organisationen lässt sich feststellen, dass Blogs sich hauptsächlich zur Unterstützung von Diskussionsprozessen punktueller Themen eignen – und nicht etwa als ein Speicher von Informationsobjekten. Stattdessen steht ein prozessorientierter Wissensaustausch zwischen Individuen durch Sozialisierung und die Erweiterung der persönlichen Netzwerke in der Organisation im Vordergrund.[41]

Im Sinne des Wissenskreislaufs von PROBST, RAUB und ROMHARDT können Blogs somit folgende Phasen des Wissensmanagements unterstützen:

- Wissensidentifikation: Die in Blogs behandelten Themen können Wissenslücken in der Organisation identifizieren, aber auch Wissen sichtbar machen, das bisher in der Organisation nicht expliziert wurde und daher „unsichtbar" war.

---

41 Ähnlich äußert sich auch Andreas Abecker vom FZI Karlsruhe in der Süddeutschen Zeitung, Nr. 201, S.V2/12 (2007) im Artikel „Jeder kann mitmachen – Was Wikis, Blogs und andere „Social Software" für Wissenstransfer und Personalentwicklung leisten".

- Wissensverteilung: Durch die Ausrichtung auf eine möglichst hohe Nutzerzahl werden Informationen einem breiten Personenkreis zugänglich gemacht.
- Wissensentwicklung: Durch Diskussion von Themen in Blogs kann das explizierte Wissen erweitert und rekombiniert werden. Dabei bleibt der Effekt jedoch durch die themenbezogene Fokussierung der Blogs vermutlich begrenzt.

Als kritische Erfolgsfaktoren für einen nutzbringenden Einsatz von Blogs im Sinne eines durchdachten Wissensmanagementkonzeptes werden vermutet:

- Der Einbezug von führenden Fachexperten und des Managements als Blogger, da der Informationsgehalt eines Blogs in hohem Maße von den tätigen Persönlichkeiten und der Qualität und Relevanz der Inhalte abhängt. Die Qualität von Blogs ist besonders in Unternehmen besonders entscheidend, da das Lesen und Kommentieren von Blogeinträgen mit dem Tagesgeschäft um die zur Verfügung stehende Arbeitszeit konkurriert.
- Die Schaffung einer wissensfreundlichen Unternehmenskultur, wie sie auch im Rahmen anderer Wissensmanagementinitiativen propagiert wird.
- Die institutionalisierte Nutzung von Blogs zur Abwicklung bestimmter wiederkehrender Prozesse (z. B. des betrieblichen Vorschlagswesens oder zur Diskussion und Ergebnisverbreitung über die Ergebnisse von Besprechungen, Workshops oder anderer aktueller Themen). So wird eine breite Nutzerbasis im Unternehmen auf die Blogs aufmerksam gemacht. Dies ist wichtig, da ein positiver Nutzenzusammenhang zwischen der Zahl der beteiligten Personen und der Anzahl der Beiträge in den Blogs unterstellt wird.
- Für weitreichende Informationen sollten Blogs auf weiterführende Informationsartefakte (z. B. Dokumente) in der Organisation verlinken, damit einerseits ihre Aussagekraft erhöht wird und andererseits eine Verbindung zu Systemen des dokumentenorientierten Wissensmanagements hergestellt wird (Vermeidung isolierter Wissensinseln).
- Kombinierung eines *Top-down-* (von der Unternehmensführung propagierte, gesteuerte Inhalte) mit einem *Bottom-up*-Managementkonzept (*User-Generated Content* der Arbeitnehmer in unreglementierterweise). So können sowohl Blogs mit aktuell wichtigen Themen im Unternehmen als auch private Blogs mit einer hohen Bandbreite an Themen unterstützt werden.

4 Zusammenfassung und Ausblick

Wissensmanagement ist ein Managementkonzept zum „guten“ Umgang mit der Ressource Wissen in Unternehmen. Auf der anderen Seite stellen Corporate Blogs spezielle Ausprägungen von Blogs im Sinne des Web 2.0 dar. Sie sind als Webseiten mit zeitlich geordneten Einträgen zu verstehen und leisten einen technischen und sozialen Beitrag zur Schaffung und Verbreitung von Wissen in Organisationen.

Bezogen auf die Phasen der Wissensschaffung in Organisationen (Sozialisation, Externalisierung, Kombination, Internalisierung) eigen sich Corporate Blogs hauptsächlich zur diskussionsorientierten Verbreitung und Erweiterung von punktuellem Wissen. In diesem Zuge können Corporate Blogs eine Erweiterung traditioneller Wissensmanagementkonzepte, die sich häufig

mit der Organisation von Informationsartefakten (Informationsmanagement) oder mit der Wissensverbreitung in Arbeitsgruppen (Projektteam, Qualitätszirkel) befassen, darstellen.

Dabei werden insbesondere die Wissensmanagementphasen Wissensidentifikation, Wissensverteilung und Wissenserweiterung unterstützt. Bezogen auf die anderen Phasen bleibt der Effekt aufgrund der Eigenschaften von Blogs jedoch begrenzt.

Für einen sinnvollen Einsatz von Blogs in Organisationen werden Maßnahmen zur Erzielung einer möglichst hohen Aussagekraft der Einträge, einer Vernetzung weiterführender Informationsobjekte in der Organisation und eine Institutionalisierung des Blogging in der Organisation als kritische Erfolgsfaktoren vermutet. Dies sollte im realen Einsatzumfeld anhand von Case Studies verifiziert und hinterfragt werden.

Literaturverzeichnis

Alby, T. (2007): Web 2.0 – Konzepte, Anwendungen, Technologien, 2. Aufl., München 2007.

Argyris, C. (1993): On Organizational Learning. 2. Aufl., Cambridge Massachusetts, Oxford 1993.

Argyris, C., Schön, D. A. (1978): Organizational Learning: A Theory of Action Perspective, Reading Massachusetts 1978.

Bennet, A., Bennet, D. (2003): The Partnership between Organizational Learning and Knowledge Management, in: Holsapple, C. W. (Hrsg.) (2003): Handbook on Knowledge Management Vol. 1: Knowledge Matters, Berlin et al. 2003. S. 439-455.

Cayzer, S. (2004): Semantic Blogging and Decentralized Knowledge Management, in: Communications of the ACM, 47. Jg., 2004, H. 12, S. 47-52.

Clyde, L. A. (2004): Weblogs – are you serious? The Electronic Library, 22. Jg., 2004, H. 5, S. 390-392.

Cyganski, P. (2007): Soziale Netzwerke im Web 2.0 – Chancen, Risiken und Veränderungen für Organisationen, in: Becker, J., Knackstedt, R. (Hrsg.). Wertschöpfungsnetze. Im Erscheinen, Berlin et al. 2007.

DIN (2006): PAS 1063 – Einführung von Wissensmanagement in Netzwerken. 2006.

HMD (2007). Glossar zu Web 2.0, in: Beck, A., Mörike, M., Sauerburger, H. (Hrsg.): HMD Praxis der Wirtschaftsinformatik, H. 255, Juni 2007. Download: http://hmd.dpunkt.de/glossar/glossar_255.html#weblog, letzter Zugriff: 20.08.2007.

Holsapple, C. W. (2003): Knowledge and Its Attributes, in: Holsapple, C. W. (Hrsg.) (2003): Handbook on Knowledge Management Vol.1: Knowledge Matters, Berlin et al. 2003, S. 165-188.

Jackson, A., Yates, J., Orlikowski, W. (2007): Corporate Blogging: Building community through persistent digital talk, in: Proceedings of the 40th Hawaii International Conference on System Sciences. 2007.

Krcmar, H. (2005): Informationsmanagement. 4. Aufl., Berlin et al. 2005.

Lee, S., Hwang, T., Lee, H.-H. (2006): Corporate blogging strategies of the Fortune 500 companies, in: Management Decision, 44. Jg., 2006, H. 3, S. 316-334.

Nonaka,I., Takeuchi, H. (1995): The Knowledge-Creating Company – How Japanese Companies Create the Dynamics of Innovation, Oxford, New York 1995.

Probst, G., Raub, S., Romhardt, K. (2003): Wissen managen – Wie Unternehmen ihre wertvollste Ressource optimal nutzen, 4., überarb. Aufl., Wiesbaden 2003, im WWW unter: http://wwwai.wu-wien.ac.at/~kaiser/seiw/Probst-Artikel.pdf

Riempp, G. (2004): Integrierte Wissensmanagementsysteme – Architektur und praktische Anwendung, Berlin et al. 2004.

Spiegel (2007): Mit Meinung, Mut und Modem, in: Spiegel Sonderheft: Leben 2.0 – Wir sind das Netz – Wie das neue Internet die Gesellschaft verändert, S. 30-31.

Therwall, M., Hasler, L. (2007): Blog Search Engines, in: Online Information Review, 31. Jg., 2007, H. 4, S. 467-479.

IV Technische Aspekte

Akamaiisierung von Applikationen

Ingo Düppe

1 Einführung

Eine Verteilung der Inhalte ist eine der zentralen Voraussetzungen für den Siegeszug des Internets. Aufgrund der rasant steigenden Nutzerzahl sowie der fehlenden zentralen Koordination [Pen03] ergeben sich bei der Verteilung von Inhalten im Internet erhebliche strukturelle Performanzprobleme. Dabei entpuppen sich insbesondere die Übergangspunkte zwischen den Teilnetzen des Internets (Peering- oder Routing Points) für Datenverkehr als Engpässe. Die Folge ist, dass bestimmte Kriterien der Übertragungsqualität nicht mehr gewährleistet werden können [RS03]. Kriterien wie Latenzzeiten beim Abruf von Inhalten, die zur Verfügung stehende Bandbreite sowie die Zuverlässigkeit von Verbindungspfaden zum Endbenutzer können von Inhaltsanbietern nicht beeinflusst werden. Gerade kommerzielle Inhaltsanbieter stehen somit vor der Herausforderung die entstehende Lücke zwischen Kundenwünschen und technischen Gegebenheiten zu schließen.

Traditionelle Lösungsansätze basieren auf der Replikation und Vorratsspeicherung der Inhalte sowie deren strategischer Verteilung im Internet [DC+03]. Hierbei liefern Content Delivery Networks (CDN) [Pen03] eine Infrastruktur zur Verteilung und Auslieferung der statischen Inhalte einer Webanwendung. Die *Akamai Technologie Inc.* ist dabei mit einem Anteil von 70 % [VP03] Marktführer. Sie erzielte mit der Verteilung von Inhalten im Internet im 2. Quartal 2007 einen Umsatz von 157 Mio. $. Das CDN von Akamai mit über 25.000 Servern in 69 Ländern und in nahezu 900 Netzwerken ist von 85 % der weltweiten Internetnutzer mit nur einem einzelnen Teilnetzwechsel erreichbar [Ak06a]. Zeitweise werden 20 % des gesamten weltweiten Internetdatenverkehrs über Akamai-Server abgewickelt [Ak06c].

Die traditionellen Methoden zur Skalierung der Inhaltsdistribution [Ak06b] scheitern jedoch vor dem Hintergrund von Web 2.0-Anwendungen [VH07]. Durch den wachsenden Anteil an dynamischen Inhalten, wie *User-Generated Content* oder auf Ajax basierende Rich Internet Application (RIA), sinkt die Möglichkeit der statischen Replizierung der Inhalte [Ak06b]. Ein Lösungsansatz ist die Verteilung der Anwendung innerhalb des CDN, anstatt der generierten Inhalte. Dieses Phänomen wird unter dem Begriff „Akamaiisierung von Applikationen" thematisiert, wobei auch neue Gestaltungsanforderungen an Applikationen zu stellen sind.

In Kapitel 2 werden zunächst die Herausforderungen der globalen Distribution von stark frequentierten Inhalten aufgezeigt und ein möglicher Lösungsansatz der traditionellen Content Delivery Networks vorgestellt. Im Anschluss daran werden in Kapitel 3 die Konsequenzen, die sich aus Web 2.0-Anwendungen für CDN ergeben, sowie mögliche Lösungsansätze diskutiert. Abschließend werden in Kapitel 4 eine kurze Zusammenfassung und ein Ausblick auf weitere Fragestellungen gegeben.

2 Content Delivery und Content Delivery Networks

Mit der zunehmenden Nutzung des Internets für die Abwicklung von B2B- sowie B2C-Transaktionen hat sich auch eine Reihe von neuen Herausforderungen für die Verteilung von Inhalten ergeben. Zu nennen sind hierbei insbesondere die im Folgenden dargestellten Herausforderungen, die es zu lösen gilt.

Die durch die Nutzung der Internetanwendung generierte Last kann die Leistungsfähigkeit eines Servers überschreiten. Abhilfe kann hierbei eine Serverfarm (engl. Cluster) schaffen, die aus einem Verbund von einzelnen Servern besteht. Die Serverfarm erscheint gegenüber den Nutzern (Clients) als ein einzelner leistungsstarker Server (vgl. [DC+03]). Anfragen der Nutzer werden über entsprechende Soft- oder Hardwarekomponenten transparent an einen Serverknoten der Serverfarm weitergeleitet. Somit wird die Lastverteilung zwischen den einzelnen Servern ermöglicht und die Skalierbarkeit des Dienstes insgesamt erhöht. Obwohl Serverfarmen zwar eine Skalierbarkeit des Originalservers ermöglichen, bieten sie keine Lösung bei Überlastsituationen in den Netzwerkkomponenten entlang des Verbindungspfads im Internet [Ak04b].

Ein weiterer Nachteil von Serverfarmen ist die unzureichende Flexibilität des Leistungsangebots. Um einen kontinuierlichen und unterbrechungsfreien Betrieb auch bei Lastspitzen zu ermöglichen, ist die Gesamtkapazität an der zur Bewältigung der maximalen Lastspitze notwendigen Kapazität auszurichten. In weniger lastintensiven Zeiten stehen somit ungenutzte Kapazitäten bereit, die einen erheblichen Kostenfaktor darstellen [He00]. Weitere Kapazitätsplanungsprobleme ergeben sich aus der notwendigen Prognose zukünftiger Lastspitzen [Ak04c].

Ebenfalls Probleme bereitet der Distribution von Inhalten die globale Verteilung der Nutzer. Diese können durch den alleinigen Einsatz von Serverfarmen nicht gelöst werden. In Abhängigkeit der Entfernung des Nutzers zum geografischen Standort der Serverfarm haben Nutzer mit einer besseren oder schlechteren Qualität der Verbindung hinsichtlich ihrer Verfügbarkeit und Zuverlässigkeit zu rechnen. Dabei kann sich die Qualität der Verbindungen durch plötzlich auftretende Überlastungen in einem der Transitnetze entlang des Verbindungspfads verschlechtern. Als maßgebliche Ursache dieses Problems ist das *Border Gateway Protocol* [St99] zu sehen, das nur bedingt flexibel auf solche unvorhergesehenen Engpässe reagiert [RS03].

Bei den genannten Herausforderungen liefern derzeit CDN eine sehr gute Unterstützung. Dabei ist ein CDN eine verteilte Infrastruktur zur effizienten Auslieferung von Inhalten an Clients. Diese besteht aus einer Menge von Serverknoten (Surrogate), an denen replizierte Inhalte gespeichert werden [DC+03, RS02, Bu05]. Die statischen Inhalte einer Webanwendung werden innerhalb des CDN auf die Surrogate repliziert. Diese befinden sich in geografischer (topologischer) Nähe zum Endnutzer [Pen03]. Durch die damit verbundene Verkürzung der Verbindungspfade wird eine Qualitätssteigerung der Übertragung erzielt. Im Gegensatz zu Web-Caching-Verfahren bei Proxy-Servern können CDN auch zur Replizierung von personalisierten Daten genutzt werden [Ak06b].

Die Architektur von CDN besteht im Kern aus fünf Komponenten (vgl. Abbildung 1). Der *Originalserver* beinhaltet die Webanwendung, welche die zu replizierenden Inhalte an das *Verteilungssystem* (engl. Distribution System) übergibt. Das *Verteilungssystem* übernimmt die Replizierung der Inhalte auf die Surrogate. Die Anfragen der Clients werden vom *Anfrageweiterleitungssystem* (engl. Request Routing System) bearbeitet und anschließend an ein aus Sicht des

CDN optimales *Surrogat* weitergeleitet. Die Übertragung der Inhalte wird durch das *Auslieferungssystem* (engl. Delivery System) durchgeführt. Bei Echtzeitübertragungen (*Streaming*) von Medienströmen liegt gerade im *Auslieferungssystem* ein signifikanter Teil der Funktionalität des CDN [Bu05]. Das *Monitoringsystem* ermittelt fortlaufend den Systemzustand, der zur Steuerung der Funktionalität des CDN genutzt wird. Letztlich werden alle Aktivitäten im *Accountingsystem* gesammelt und aufbereitet, sodass sie dann von der Rechnungslegung genutzt werden können.

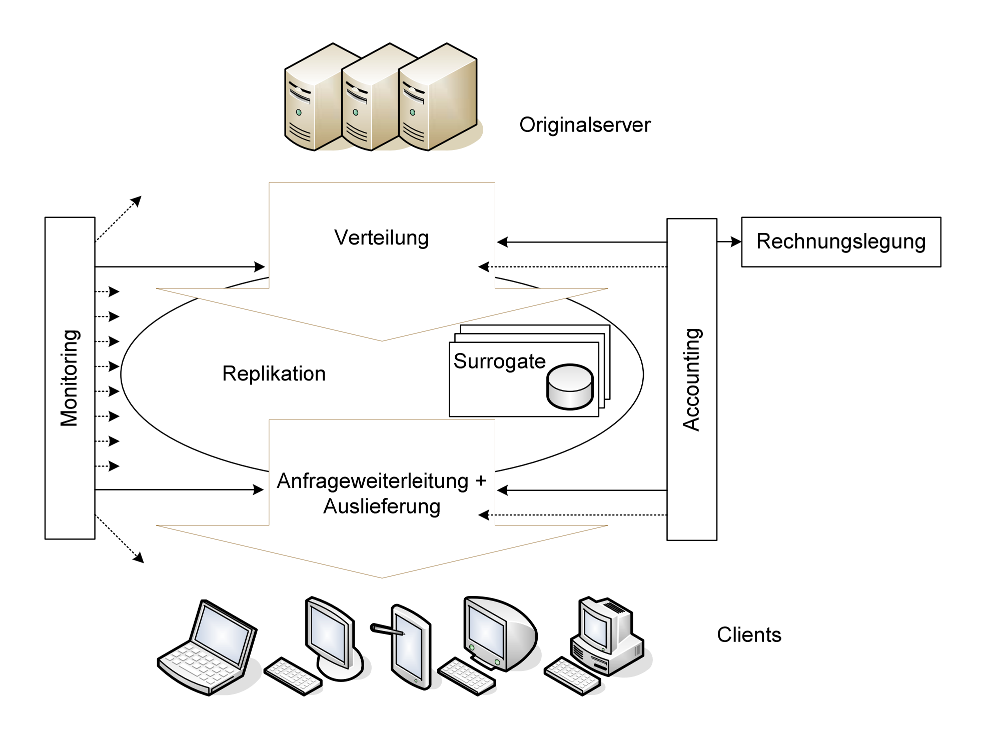


Abbildung 1: CDN-Systemkomponenten [Bu05]

Zentrale Bedeutung für CDN haben Algorithmen zur optimalen Platzierung der Replikate auf die vorhandenen Surrogate. Zwar ermöglichen dynamische Verfahren eine kontinuierliche Replatzierung der Replikate, um flexibel auf Änderungen des Systemzustands zu reagieren. Jedoch können dynamische Verfahren aufgrund der Größe des Gesamtsystems und der damit verbundenen Anzahl von Parametern nur eine lokale Optimierung vornehmen. Hingegen liefern statische Verfahren tendenziell bessere Platzierungen, können aber aufgrund ihrer Komplexität nicht kontinuierlich durchgeführt werden [SRS00].

Neben der Platzierung der Replikate ist die Bestimmung der Surrogate zur Beantwortung von Client-Anfragen für das Request-Routing-System von Bedeutung. Mögliche Kriterien hierbei sind die geografische Nähe, die Auslastung des Surrogates und die Auslastung des Netzwerks. Erstere können durch relativ einfache Metriken bestimmt werden. Die Netzwerkauslastung ist dagegen schwerer zu bestimmen. Dabei sind passive und aktive Verfahren zu unterscheiden. Bei passiven Verfahren analysieren die einzelnen Surrogate die bestehenden Verbindungen zu den Clients und übermitteln die entsprechende Metriken an das Request-Response-System. Aber diese Verfahren ermöglichen lediglich eine Ex-post-Betrachtung des Systemzustands und entsprechende Rückschlüsse auf den zukünftigen Systemzustand sind ungenau. Aktive Verfahren hingegen führen kontinuierliche Messungen durch. Zur Messung werden beispielsweise ICMP-Echo-Nachrichten versandt. Der Nachteil hierbei liegt in der zusätzlich erzeugten Last auf

das System und der Blockierung einzelner Systemkomponenten, wie Router und Firewalls, von ICMP-Echo-Abfragen.

Der transparente Zugriff des Clients auf die replizierten Inhalte des CDN kann durch verschiedene *Request-Routing-Verfahren* ermöglicht werden, die auch eine Lastverteilung realisieren. Die einfachste Variante ist ein auf dem *Domain Naming System (DNS)* [DC+03] basierendes Verfahren, bei dem der DNS-Server des CDN die IP-Adresse des optimalen Surrogat-Servers liefert. Ein DNS-basiertes *Request-Routing-Verfahren* gestattet nur eine sehr grobe Lastverteilung, da die aufgrund der Time to Live (TTL)-Direktive als temporär markierten DNS-Antworten oftmals von clientseitigen DNS-Server längerfristig gespeichert werden. Ein *Request-Routing-Verfahren*, das auf der OSI-Transportschicht [SGG00] basiert, bietet im Gegensatz den Vorteil einer effizienteren Lastverteilung. Die Kommunikation vom Client zum CDN (Upstream) erfolgt immer über das Request-Routing-System, lediglich die Daten vom CDN zum Client (Downstream) werden direkt übertragen. Das *Request-Routing-System* verfügt somit über detailliertere Informationen, wie beispielsweise die IP-Adresse und die Portnummer des Clients, und kann somit eine Lastverteilung auf Endbenutzerebene realisieren. Wird das *Request-Routing-Verfahren* auf der OSI Anwendungsschicht angesiedelt, kann eine ressourcenspezifische Lastverteilung erreicht werden. Mittels HTTP-302-Redirection [FG+99] können nur vollständige HTTP-Anfragen auf Surrogate umgeleitet werden. Hingegen können mit URL-Rewriting die Ressourcenverweise innerhalb von HTML-Seiten auf andere Surrogate umgeleitet werden. Zudem ist ein Umweg über das *Request-Routing-System* nicht mehr notwendig.

Neben der verwendeten Technik ist die Ausdehnung des CDN eine entscheidungsrelevante Größe. So werden CDN-Anbieter gemäß ihrer Abhängigkeit zu Internet-Service-Providern (ISP) unterschieden. Befinden sich sämtliche Surrogate des CDN innerhalb eines einzelnen ISP, so wird es als *single-ISP* bezeichnet (z. B. AT&T). Entsprechend werden CDN, deren Surrogate sich über mehrere ISP-Netze erstrecken, als *multi-ISP* bezeichnet [RS03]. Das von Akamai Technology betriebene CDN ist ein multi-ISP [Buc05].

Zur Auswahl eines Surrogates können neben technischen Parametern auch ökonomische Kriterien herangezogen werden. Beispielsweise kann nach dem kostengünstigsten Surrogat gesucht werden, der eine bestimmte Verbindungsqualität erfüllt.

3 Konsequenzen von Web 2.0 für Content Delivery Networks

Aus Sicht des Web 2.0 [Be07] verschärfen sich die oben genannten Herausforderungen weiter; gleichfalls treten neue Herausforderungen auf. Die *Ajax*-basierten Rich Internet Applicationen benötigen meist geöffnete TCP-Verbindungen und ihre Interaktionen mit der Webanwendung sind deutlich häufiger [Bo07]. Ebenso sinkt der Anteil von statischen zu dynamischen Inhalten von Webanwendungen, während der Anteil von *User-Generated Content* in heutigen Anwendungen steigt [DM+02]. Zusammen führt dies zu einer verstärkten Kommunikation zwischen dem Client (Endnutzer) und der Webanwendung sowie dem dahinter liegendem Backend-System.

Eine Lösung dieser Herausforderungen durch eine einfache Anwendung des CDN-Prinzips im Web 2.0 erscheint wenig erfolgversprechend. Vielmehr sind neue Gestaltungsprinzipien für die

Verteilung von Inhalten zu finden. Potenzielle Lösungen für die Verteilung von Inhalten im Web 2.0-Kontext sind *Edge Side Include* [TW+01] sowie *Edge Computing* [Ak03].

Um dem Phänomen steigender Anteile von dynamischen Elementen in Webseiten entgegenzuwirken, wurde auf Initiative von Akamai Technologies Inc. *Edge Side Include* beim World Wide Web Consortium spezifiziert. Diese XML-basierte Sprache ermöglicht es, dynamische Seiten nicht mehr als Ganzes betrachten zu müssen, sondern in einzelne Fragmentelemente zu unterteilen. Bei jedem Fragment kann die Gültigkeitsdauer definiert werden, bevor der Surrogate bei erneuter Client-Anfrage eine aktualisierte Version vom Originalserver anfordern muss (vgl. Abbildung 2) [Ak04a].

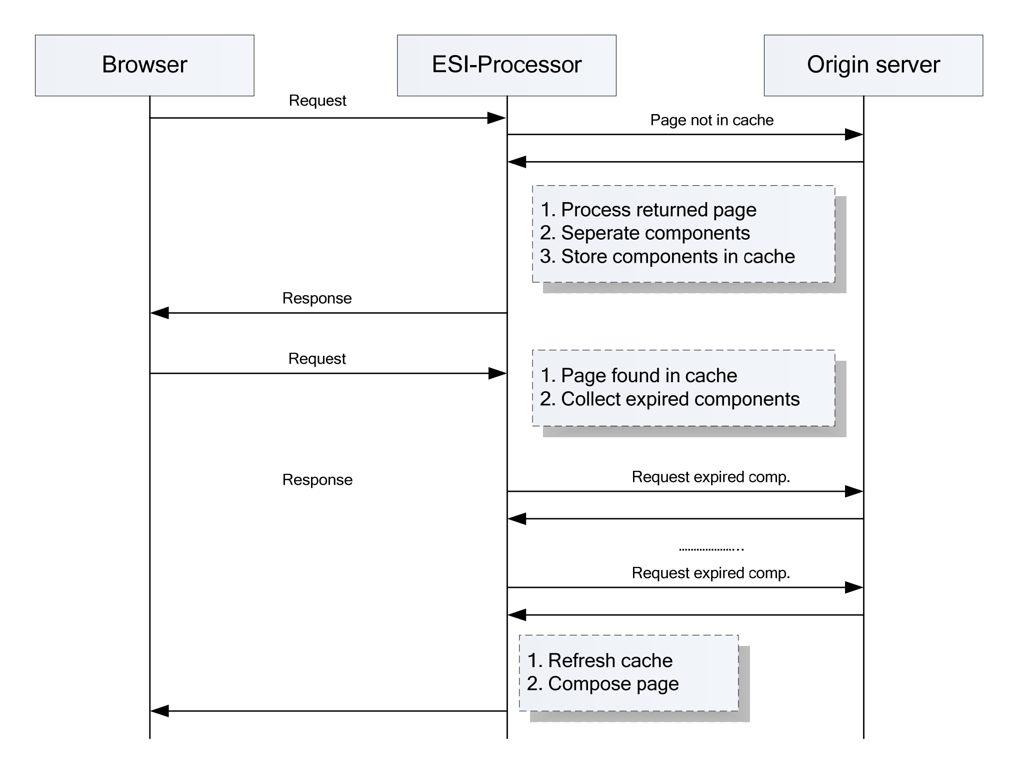


Abbildung 2: Edge Side Include-Anfrageprozess [KM02]

Die Nachteile von ESI sind identisch den bekannten Problemen von Proxy-Servern. Da Aktualisierungen nicht aktiv in die Surrogate propagiert werden können, besteht die Möglichkeit, dass dem Endnutzer inkonsistente oder veraltete Daten präsentiert werden.

Beim *Edge Computing*, eine spezielle Form des Grid Computings [Su03], wird hingegen versucht diese Nachteile zu umgehen, indem die Webanwendung auf die *Edge Computer* (Surrogate) des CDN verteilt wird. Akamai Technologies bietet dazu seine *Edge Computing-Plattform* auf Basis des Java EE Websphere Applikationserver von IBM an [*Ak04b*] (vgl. Abbildung 3).

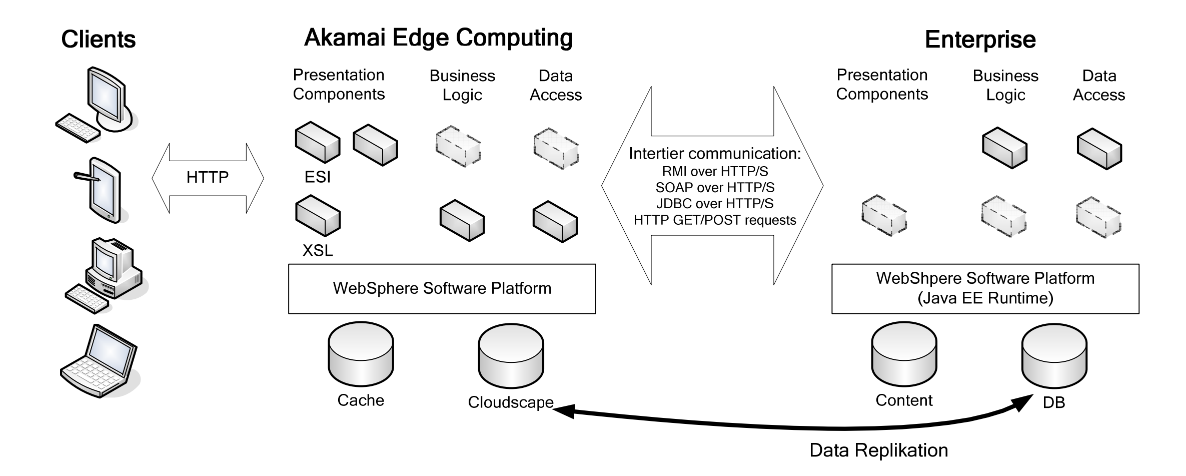


Abbildung 3: Edge Computing-Plattform [Ak04d]

Edge Computing geht von Anwendungen aus, die nach den Gestaltungsprinzipien der Java Enterprise Edition [Su99] aus Präsentations-, Geschäftslogik- und Datenzugriffskomponenten bestehen. Dabei werden die existierenden Komponenten aus dem Originalserver herausgelöst und auf die Surrogate übertragen. Die damit verfolgten Skalierungseffekte entsprechen zunächst den der Serverfarmen, profitieren aber zusätzlich von der weltweiten Verteilung [Ta04]. Durch die Verteilung der Präsentationskomponenten und der Nähe zu den Clients stellen somit auch Ajax-Verbindungen keine Performanzprobleme dar. Vielmehr wird die Herausforderung der Skalierung bei einer solchen Architektur auf die Kommunikation zwischen den Komponenten auf dem *Edge Computer* und den auf dem Originalserver verschoben. Zwar bietet *Edge Computing* die Möglichkeit, die Kommunikationsprotokolle mittels Webservices oder HTTP transparent zu kapseln, doch die drastische Reduzierung des Datenverkehrs zwischen den jeweiligen Komponenten obliegt nach wie vor der Anwendung. Hierbei sollen intelligente Caching-Mechanismen helfen, den Datenverkehr zwischen *Edge Computer* und *Originalserver* zu puffern. Beispielsweise kann bei einem Onlineshop der Produktkatalog auf den *Edge Computer* gepuffert werden; lediglich Informationen über die Verfügbarkeit einzelner Produkte müssen vom Originalserver abgefragt werden [LK+07].

4 Zusammenfassung und Ausblick

Durch eine zunehmende kommerzielle Nutzung des Internets ergeben sich für die Verteilung von Inhalten neue Herausforderungen. Aufgrund von Performanzproblemen ist eine einfache Replikation der Inhalte unzureichend. Als Lösungsansatz hat sich das Prinzip der so genannten *Content Delivery Networks* herausgebildet, wodurch eine performante Verteilung von Inhalten ermöglicht wird. Durch den Trend, Inhalte zunehmend dynamisch zu gestalten, wie es vor allem im Bereich von Web 2.0-Anwendungen zu beobachten ist, stoßen diese Gestaltungsprinzipien an ihre Grenzen. Aus diesem Grund sind neue Konzepte zu entwickeln, die auch eine Verteilung dynamischer Inhalte unterstützten. Besondere Potenziale werden hierbei dem *Edge Side Include* sowie dem *Edge Computing* zugerechnet. Letzteres ist dem *Edge Side Include* vorzuziehen, da hierbei einwertige Aussagen über eine Aktualisierung des Cache der Surrogate ermöglicht werden.

Bislang weitestgehend ungelöste Fragestellungen, die durch die Nutzung dieser Prinzipien für Web 2.0-Anwendungen hervorgerufen werden, sind u. a. das transparente Cachen der Kommunikation zwischen den Komponenten der Anwendung auf dem *Edge Computer* und den Komponenten auf dem Originalserver sowie die erforderliche Trennung der Inhaltsfragmente von Web 2.0-Anwendungen in statische und dynamische.

Ebenfalls ungelöst sind Aspekte der Gestaltung von Verrechnungspreisen und Preismodellen für die Verteilung von Web 2.0-Inhalten innerhalb von CDN. Hierfür sind geeignete Controlling-instrumente zu entwickeln, die eine verursachungsgerechte Zurechnung von Kosten ermöglichen und damit eine differenzierte Marktbearbeitung unterstützen.

Literaturverzeichnis

[Ak04a] Akamai Technologies: EdgeSuite 5.0 – ESI Developer's Guide. http://www.akamai.com/dl/technical_publications/akamai_esi_developers_guide.pdf [07.09.07], 2004.

[Ak04b] Akamai Technologies: Akamai Edge Computing. Enabling Applications that grow your business. http://www.akamai.com/dl/whitepapers/Akamai_Enabling _Apps_Grow_Business_Whitepaper.pdf [07.09.07], 2004.

[Ak04c] Akamai Technologies: The Business Internet – A predictable platform for profitable e-business. http://www.akamai.com/dl/whitepapers/Akamai _Business_Internet_Whitepaper.pdf [07.09.07], 2004.

[Ak04d] Akamai Technologies: A Developer's Guide to On-Demand Distributed Computing. Best Practives for Design and Deployment of High-Performance J2EE Applications. http://www.akamai.com/dl/whitepapers/Akamai_DeveloperGuide_Distributed_Computing.pdf [07.09.07], 2004.

[Ak06a] Akamai Technologies: Facts & Figures. http://www.akamai.com/html/about/facts_figures.html [07.09.07], 2006.

[Ak06b] Akamai Technologies: Web 2.0 is Here – Is Your Web Infrastructure Ready. http://www.akamai.com/html/forms/web20_postcard.html [07.09.07], 2006.

[Ak06c] Akamai Technologies: Akamai ermöglicht einmaligen Einblick ins aktuelle Internet-Geschehen. http://www.akamai.de/dl/PM_Akamai_Data_Visualization _Tool_190607.pdf [07.09.07], 2006.

[Be07] Beck, A.: Web 2.0 – Konzepte, Technologie, Anwendung, in: HDM, Praxis der Wirtschaftsinformatik, 2007, H. 255, S. 5-16.

[Bo07] Bosch, A.: Ajax – Grundlagen und Funktionsweise, in: HDM, Praxis der Wirtschaftsinformatik, 2007, H. 255, S. 37-48.

[Bu05] Buchholz, S.: Adaptivitätssensitive Platzierung von Replikaten in Adaptiven Content Distribution Networks, Dissertation, Technische Universität Dresden 2005.

[DC+03] Day, M., Cain, B., Tomlinson, G., Rzewski, P.: A Model for Content Internetworking (CDI). RFC 3466. http://rfc.net/rfc3466.html [07.09.07], 2003.

[DM+02] Dilley, J., Maggs, B., Parikh, J., Prokop, H., Sitaraman, R., Weihl, B.: Globally Distributed Content Delivery, IEEE Internet Computing, 7. Jg., 2002, H. 5, S. 50-58.

[FG+99] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., Berners-Lee, T.: Hypertext Transfer Protocol – HTTP/1.1. http://www.w3.org/Protocols/rfc2616/rfc2616.html [07.09.07], 1999.

[He00] Heap, D. G.: Scorpion – Simplifying the Corporate IT Infrastructure. http://www.ibm.com/servers/eserver/zseries/library/whitepapers/pdf/gf225168.pdf [07.09.07], 2000.

[KM04] Kaundinya, M., Muralidharan, S.: Edge Computing with Java Edge Side Includes. An end-to-end architectural illustration. Java Developer Journal, http://java.sys-con.com/read/45496.htm [07.09.07], 2004.

[LK+04] Lin, Y., Kemme, B., Patino-Martinez, M., Jiménez-Peris, R.: Enhancing Edge Computing with Databases. McGill University, School of Computer Science, Canada. http://lsd.ls.fi.upm.es/lsd/papers/2007/srds07-final.pdf [07.09.07], 2007.

[Pen03] Peng, G.: CDN: Content Distribution Network, Technischer Bericht TR-125, Experimental Computer Systems Lab, Department of Computer Science, State University of New York, New York 2003.

[RS03] Rabinovich, M., Spatscheck, O.: Web Caching and Replication, Reading, MA 2003.

[SGG00] Silberschatz, A., Galvin, P., Gagne, G.: Applied Operation System Concept, New York 2000.

[SRS05] Song, Y. J., Ramasubramanian, V., Sirer, E. G.: Optimal Resource Utilization in Content Distribution Networks. Dept. Of Comuter Science, Cornell University, Ithaca, New York 14853. http:// www.cs.cornell.edu/People/egs/papers/cobweb.pdf [07.09.07], 2005.

[St99] Stewart, J. W.: BGP4: Inter-Domain Routing in the Internet, Reading, MA 1999.

[Su99] Sun Microsystem: Simplified Guide to the Java 2 Platform, Enterprise Edition. http://java.sun.com/j2ee/reference/whitepapers/j2ee_guide.pdf [07.09.07], 1999.

[Su03] Sun Microsystems: Computing at the Edge – White Paper. http://www.sun.com/software/grid/whitepaper.edge.pdf [07.09.07], 2003.

[Ta04] Taylor, K.: How to Improve J2EE Performance and Reliability. Computerworld. http://www.computerworld.com/softwaretopics/software/appdev/story/0,10801,94202,00.html [07.09.07], 2004.

[TW+01] Tsimelzon, M., Weihl, B., Chung J., Frantz, D., Basso, J., Newton, C., Hale, M., Jacobs, L., O'Connel, C.: ESI Language Specification. W3C Note August. http://www.w3.org/ TR/esi-lang [07.09.07], 2001.

[VH07] Vossen, G, Hagemann, S.: Unleashing Web 2.0 – from concepts to creativity, Amsterdam, Bosten 2007.

[VP03] Vakali, A., Pallis, G.: Content Delivery Networks: Status and Trends, IEEE Internet Computing, 7. Jg., 2003, H. 6, S. 68-74.

IPTV

Philipp Bergener

1 Einleitung

Internet Protocol Television oder kurz IPTV beschreibt eine neue Art der Fernsehübertragung mittels des Internet Protocolls (IP). In diesem Beitrag soll untersucht werden, welche Möglichkeiten sich aus diesem neuen Verbreitungskanal ergeben. Dazu wird in Kapitel 2 zunächst definiert, was IPTV ist, was nicht mehr unter diesen Begriff fällt und welche Dienste mit IPTV möglich sind. Kapitel 3 gibt einen kurzen Überblick über die technischen Aspekte im Zusammenhang mit IPTV. Kapitel 4 zeigt Ausgestaltungsmöglichkeiten für IPTV-Geschäftsmodelle auf und stellt einige Beispiele vor. In Kapitel 5 werden einige sozio-ökonomische Konsequenzen beschrieben. Kapitel 6 gibt schließlich einen Ausblick über mögliche zukünftige Entwicklungen im Bereich IPTV.

2 Was ist IPTV

2.1 Definition und Abgrenzung

Internet Protocol Television oder kurz IPTV bezeichnet die Übertragung von digitalem Fernsehen und anderen Multimediainhalten mithilfe des Internet Protocols, welches auch zur Datenübertragung im Internet eingesetzt wird [Wik, CoLo06, S. 7]. Die internationale Fernsehunion ITU definiert IPTV folgendermaßen:

„IPTV is defined as multimedia services such as television/video/audio/text/graphics/data delivered over IP-based networks managed to support t*he required level of QoS/QoE, s*ecurity, interactivity and reliability.“ [ITU07]

Dies beinhaltet als weitere Aspekte eine gewisse Qualität der Übertragung, entweder als Dienstgüte (QoS – Quality of Service) oder subjektiv aus Sicht des Konsumenten (QoE – Quality of Experience), sowie Sicherheit, Interaktivität, d. h. einen Rückkanal vom Empfänger zum Sender, und Zuverlässigkeit. Besonders diese Interaktivität unterscheidet IPTV vom klassischen Fernsehen, das analog oder digital per Satellit, über Kabel oder terrestrisch übertragen wird.

An dieser Stelle ist es sinnvoll, IPTV gegenüber dem Begriff Internet-TV, der zum Teil synonym verwendet wird, abzugrenzen. Wird eine Unterscheidung zwischen IPTV und Internet-TV vorgenommen, so wird als häufigstes Unterscheidungskriterium das Netzwerk, über das die Übertragung erfolgt, herangezogen. Nach dieser Unterscheidung wird IPTV über geschlossene, proprietäre Netzwerke, z. B. von Telekommunikations- oder Kabelanbietern, übertragen, während Internet-TV über das öffentlich zugängliche Internet verbreitet wird. Basis für diese Unterscheidung ist, dass das Internet als Best-Effort-Dienst entsprechende Dienstgüte sicherstellen kann [Goo05].

Diese Definition wird hier allerdings als zu eng empfunden. Zum einen würde dadurch eine Reihe von interessanten Ansätzen aus dem Rahmen dieser Arbeit fallen, weil sie über das Internet übertragen werden. Zum zweiten steigt die Bandbreite der Internetzugänge kontinuierlich, sodass auch hier eine entsprechende Dienstqualität erreicht wird. Daher wird hier der Auffassung von [CoLo06, S. 8] gefolgt, dass es sich bei der Übertragung über proprietäre bzw. öffentliche Netze vielmehr um zwei verschiedene Geschäftsmodelle des IPTV handelt. Um jedoch nicht jedes Video im Internet unter IPTV zu fassen, wird hier der Definition von [Gug06] gefolgt, nach der bei IPTV die Übertragung selektiv erfolgt, d. h., der Anbieter behält eine gewisse Kontrolle über den Inhalt. Internet-TV dagegen „umfasst speziell für das Internet aufbereitete Videos und Filme, die über Webseiten angeboten werden". Damit zählen insbesondere Seiten wie YouTube nicht zum IPTV.

2.2 Dienste

Beim klassischen Fernsehen, das über Kabel, Satellit oder terrestrisch ausgestrahlt wird, handelt es sich um programmorientiertes, lineares Fernsehen. Dies bedeutet, dass das Fernsehprogramm in einer festen Folge zu vordefinierten Zeitpunkten für die einzelnen Sendungen ausgestrahlt wird. Der Konsument kann lediglich auswählen, welchen Kanal er sehen möchte, muss sich aber ansonsten mit den zu einem Zeitpunkt laufenden Sendungen begnügen. Auch über IPTV ist diese Art des linearen Fernsehens möglich. Allerdings erlaubt es auch weitere Formen und zusätzliche Dienste:

So ist es möglich, einen bestimmten Inhalt, z. B. einen Film, zu einem beliebigen Zeitpunkt abzurufen. Dies wird als Video-on-Demand bezeichnet. Sind (mehrere) feste Startzeiten für einen Inhalt vorgegeben, spricht man von Near-Video-on-Demand. Weitere Funktionen die mittels IPTV zur Verfügung gestellt werden können, sind z. B. ein persönlicher Videorekorder, Mehrkanalton oder eine elektronische Programmzeitschrift [Wik].

3 Technische Aspekte

Bei IPTV wird das Programm digital übertragen. Dazu muss das Fernsehsignal oder der Videoinhalt codiert werden. Hierzu werden Standards wie MPEG2 oder der neuere H.264/MPEG4 verwendet. Zum Empfang in Fernsehqualität wird eine Datenrate zwischen 2 bis 6 Mbit/s notwendig, für hochauflösendes Fernsehen (HDTV) 6 bis 16 Mbit/s [Wik].

IPTV benötigt im Gegensatz zum klassischen Fernsehen andere Übertragungsarten. Letzteres wird klassischerweise im *Broadcasting*-Verfahren übertragen. Es wird dabei von einem zentralen Punkt an eine unbestimmte Anzahl von Teilnehmern übertragen. Es handelt sich dabei um ein Push-Verfahren, da das Programm unabhängig von einer konkreten Anforderung des Empfängers gesendet wird. Man spricht in diesem Zusammenhang auch von Rundfunk. Die Anzahl der Teilnehmer ist dabei beliebig groß; durch den Empfang tritt kein Verbrauch des gesendeten Signals auf [Kar06, S. 30 f.].

Bei IPTV ist eine solche Art der Übertragung nicht möglich. Die hier am häufigsten vorzufindende Übertragungsart ist der *Unicast*. Hier wird jedem Empfänger einzeln das Programm vom Sender übermittelt, wie es z. B. auch beim Aufruf einer Internetseite der Fall ist. Dies hat jedoch

für den Sender u. U. eine starke Beanspruchung seiner Bandbreite zur Folge. Würde eine beliebte Sendung mit einer Datenrate von 2 Mbit/s beispielsweise von 100.000 Menschen gleichzeitig abgerufen, müsste der Sender eine Bandbreite von 200 Gbit/s zur Verfügung haben [CoLo06, S. 29; Kar06, S. 31].

Eine Alternative stellt hier *Multicast*, zu Deutsch auch Gruppenruf, dar. Hier wird vom Sender eine bestimmte Gruppe von Empfängern angesprochen. Gleicher Inhalt für mehrere Empfänger muss hier nicht mehrfach gesendet werden. Allerdings setzt dies voraus, dass alle Empfänger gleichzeitig das gleiche Programm sehen wollen, sodass dieses Verfahren für einige IPTV-Dienste, wie z. B. Video-on-Demand, nicht verwendet werden kann. Nur ein Near-Video-on-Demand mit festgelegten Startzeiten ist dann möglich. Weiterhin wird ein IP-Multicast meist nur in proprietären Netzwerken eingesetzt, da hierfür spezielle Hardware vonnöten ist [CoLo06, S. 38 ff.; Kar06, S. 31].

Eine weitere Möglichkeit, die Last des Senders zu reduzieren, bilden Peer-To-Peer-Technologien. Hier wird jeder Empfänger auch zum Sender, indem er die empfangenden Daten (für einen gewissen Zeitraum) speichert und anderen Empfängern zur Verfügung stellt. Damit dies funktioniert, ist allerdings eine gewisse Mindestanzahl von Benutzern notwendig [CoLo06, S. 28, Kar06, S. 91].

4 Geschäftsmodelle

4.1 Ausgestaltungsmöglichkeiten

Im Zusammenhang mit IPTV gibt es verschiedene Geschäftsmodelle, die sich in verschiedenen Ausgestaltungsdimensionen unterscheiden. Die erste Dimension, die bereits erwähnt worden ist, ist das Netzwerk, über das die Übertragung erfolgt. Hier gibt es auf der einen Seite geschlossene, proprietäre Netzwerke, über die Telekommunikations- oder Kabelanbieter verfügen. Geschlossene Netze setzen natürlich Infrastrukturinvestitionen aufseiten der Netzbetreiber voraus. Dafür können sie eine bestimmte Dienstgüte garantieren (z. B. durch Techniken wie Multicast). Auf der anderen Seite kann das Internet als Übertragungskanal benutzt werden. Hier kann eine bestimmte Dienstgüte aufgrund der Konstruktion des Internets als Best-Effort-Dienst, bei dem ein Datenverlust durchaus möglich ist, nicht garantiert werden. Allerdings wird die Qualität aufgrund der zunehmenden Bandbreite der Internetanschlüsse im besser [CoLo06, S. 7 ff u. S. 37].

Eng an den Übertragungsweg gekoppelt ist die Frage, ob das Programm mittels eines Computers oder über einen Fernseher gekoppelt mit einer Settopbox empfangen wird. Typischerweise werden über das Internet übertragende Angebote mit einem Computer und über proprietäre Netzwerke gesendete Inhalte mit einer Settopbox empfangen [CoLo06, S. 7 ff.].

Weiterhin können IPTV-Angebote anhand der verfügbaren Inhalte differenziert werden. Dabei kann zum einen die Form unterschieden werden, in der der Inhalt angeboten wird. Es lässt sich grundsätzlich zwischen Video-on-Demand und linearen Angeboten unterscheiden. Bei ersterem sind die Inhalte, wie beschrieben, zu beliebigen Zeitpunkten abrufbar. Letzteres entspricht dem klassischen Fernsehen mit festen Sendezeiten. Es ist allerdings für bestimmte Inhalte, allen

voran Live-Übertragungen, z. B. von Sportereignissen, zwingend. Zudem kann unterschieden werden, ob die Inhalte von „klassischen“ Fernsehsendern übernommen werden oder speziell für das IPTV-Angebot produziert worden sind [CoLo06, S. 7 ff, Kar06, S. 189].

Eine weitere Unterscheidung kann nach der Art der Finanzierung des Angebotes gemacht werden. Hier kann grundsätzlich zwischen werbefinanzierten Angeboten und solchen, für die der Nutzer bezahlen muss, unterschieden werden. Allerdings sind hier verschiedenste Ausgestaltungen wie Pay-per-View, Abos oder nur zum Teil freie bzw. über Werbung finanzierte Angebote denkbar [CoLo06, S. 7 ff.].

4.2 Beispiele

TVNZ ondemand: TVNZ ondemand ist ein IPTV-Angebot des staatlichen neuseeländischen Fernsehens. Auf lange Sicht möchte TVNZ dieses Angebot als dritten Kanal neben seinen beiden klassischen Fernsehkanälen etablieren. TVNZ stellt sein Angebot über das Internet zur Verfügung. Dies beinhaltet sowohl freie als auch kostenpflichtige Video-on-Demand-Inhalte. Erstere kann jeder Nutzer aus Neuseeland über einen Web-Player abspielen. Für letztere muss der Nutzer sich registrieren und ein Guthaben in Form von Punkten erwerben. Eine 30-minütige Sendung kostet vier Punkte, eine einstündige acht. Da aus den kostenpflichtigen Sendungen die Werbung entfernt wird, dauern sie real allerdings 23 bzw. 44 Minuten. Ein Punkt kann für 2 neuseeländische $ erworben werden. Die kostenpflichtigen Inhalte werden von der Seite heruntergeladen und mithilfe des Windows Media Players abgespielt. Dieser übernimmt auch die Authentifizierung des Nutzers. TVNZ ondemand zeigt bisher nur Inhalte der beiden klassischen Kanäle. Allerdings soll es in Zukunft auch exklusive Angebote für den IPTV-Kanal geben [Mac07, Tvn].

Joost: Joost ist ein IPTV-Angebot der Skype- und Kaaza-Gründer Niklas Zennström und Janus Friss. Bei Joost werden die Inhalt mithilfe von Peer-to-Peer-Technologie über das Internet übertragen. Zum Empfang ist ein spezieller Joost-Client von Nöten. Dieser kann über die Joost-Webseite (www.joost.com) gegen Registrierung bezogen werden. Zurzeit befindet sich Joost noch in der Beta-Phase, sodass der Download des Clients nur nach Einladung durch ein Joost-Mitglied möglich ist. Die Inhalte von Joost werden vollständig als On-Demand-Videos angeboten. Sie werden dabei thematisch in verschiedenen Kanälen gruppiert, wie z. B. „Comedy Central“ oder „Soccer Channel“. Die Inhalte bezieht Joost von Produzenten wie Viacom (Paramount, MTV) oder Warner Brothers. Ein Teil davon ist exklusiv auf Joost verfügbar. Finanziert wird Joost über Werbung, die ähnlich wie beim klassischen Fernsehen als Teil des Programms gezeigt oder in den Player eingeblendet wird [Cat207a, Joo].

Ähnliche Angebote wie Joost sind auch bei Babelgum (www.babelgum.com) und Zattoo (www.zattoo.com) zu finden, wobei Zattoo ausschließlich Live-Streams klassischer Fernsehkanäle zeigt [Cat2007a].

T-Home: T-Home ist ein Angebot der deutschen Telekom, das neben einem Telefon- und Internetanschluss auch ein IPTV-Angebot umfasst. Übertragen wird T-Home über das Telefonnetz der Telekom. Der Empfang erfolgt mithilfe einer Settop-Box am Fernseher. Ursprünglich war zur Nutzung des IPTV-Angebots ein VDSL-Anschluss notwendig, der nur in wenigen Ballungsräumen verfügbar ist; inzwischen ist der Empfang aber auch über ADSL2+ möglich. T-Home bietet

drei verschiedene Inhaltspakete: Das Paket „Entertain Basic“ bietet per Video-on-Demand-Zugriff auf Filme und ein TV-Archiv. „Entertain Comfort“ enthält zusätzlich noch ein Angebot von ca. 70 frei empfangbaren Sendern und bietet einen digitalen Videorekorder. „Entertain Plus“ bietet zusätzlich noch 30 „Premium“-Sender und zwei Fußballbundesligaspiele live pro Spieltag. Die Finanzierung des Angebots erfolgt zum einen über die Grundgebühr für die einzelnen Pakete; zum anderen sind die Video-on-Demand-Filme kostenpflichtig (0,99 € bis 3,99 €) [Tho].

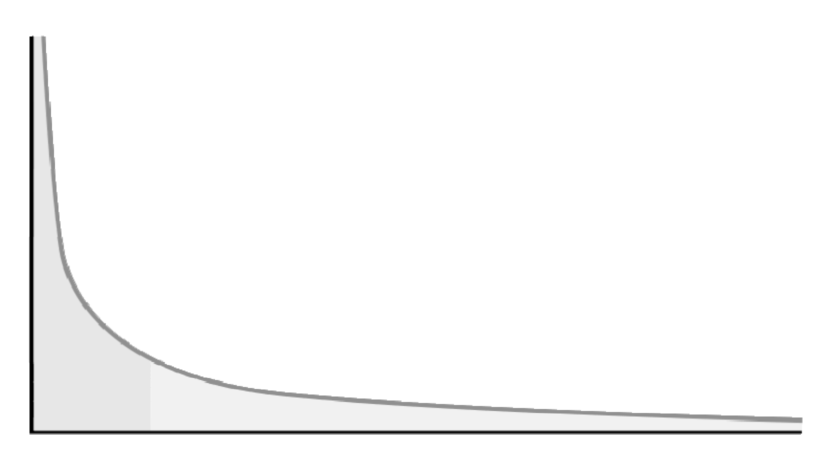
	Joost	TVNZ ondemand	T-Home
Übertragung & Empfang	Internet (P2P) Computer (Client)	Internet Computer (Webplayer, WMP)	Proprietäres Netzwerk Settop-Box
Inhalte	Video-on-Demand z. T. exklusiv	Video-on-Demand Wiederverwendung existierender Inhalte, exklusiv geplant	Video-on-Demand, lineares Programm Hauptsächlich normale Fernsehsender, Filmarchiv
Finanzierung	Werbung	Werbung, Pay-per-View	Grundgebühren, Pay-per-View

Abbildung 1: Vergleich der vorgestellten Geschäftsmodelle

Abbildung 1 fasst die vorgestellten Geschäftsmodelle noch einmal vergleichend zusammen. Es fällt dabei auf, dass es verschiedene Arten der Übertragung, des Inhalts und der Finanzierung gibt. Allerdings bieten alle Angebote Video-on-Demand in irgendeiner Form an.

5 Sozio-ökonomische Aspekte

5.1 Long-Tail-Effekt



Quelle: [Wik]

Abbildung 2: Long-Tail-Effekt (Masse: dunkelgrau, Nische: hellgrau)

Ausgangspunkt hier ist die Beobachtung, dass auf den meisten Märkten eine relativ kleine Anzahl von Produkten die größte Nachfrage nach sich zieht (Masse, Abbildung 2, dunkelgrau), während der restliche Großteil der Produkte nur eine kleine Nachfrage auf sich vereint (Nische, Abbildung 2, hellgrau). Auf das Fernsehen übertragen bedeutet dies, dass es eine kleine An-

zahl relativ populärer Programme gibt und einen Großteil von Nischenprogrammen mit einem relativ geringen Zuschauerzuspruch. Aufgrund des mangelnden Platzangebots in den klassischen Fernsehübertragungswegen wie Kabel oder Antenne und der zeitlichen Beschränkung durch das lineare Fernsehen (maximal 24 Stunden pro Tag und Kanal), werden die Nischenprogramme meist nicht gezeigt. Bei IPTV sind diese Beschränkungen nicht länger vorhanden. Theoretisch können mehrere Millionen Kanäle gleichzeitig übertragen werden. Weiterhin hebt TV-on-Demand die zeitlichen Beschränkungen des linearen Fernsehens auf. Mit IPTV können auch Nischenprogramme einfach an die entsprechenden Interessenten übertragen werden [CoLo06, S. 40 f.; And04].

5.2 Werbung

Werbung im herkömmlichen Fernsehen leidet unter einer großen Fragmentierung der Zielgruppen, die auf dem klassischen Weg nur sehr schwer gezielt angesprochen werden können, auch aufgrund der im vorigen Kapitel angesprochenen beschränkten Anzahl von Kanälen, die bemüht sind, ein breites Publikum anzusprechen. So kann die Ausrichtung der Werbung nur bedingt an den Inhalten vorgenommen werden. Zudem ist es über die herkömmlichen Verbreitungswege schwer möglich, ortsbasierte (location based) Werbeinhalte anzubieten [Chr05; CoLo06, S. 41].

IPTV bietet hier neue Möglichkeiten. Zum einen erlaubt der im vorigen Kapitel beschriebene Long-Tail-Effekt ein größeres Angebot von Inhalten, insbesondere für Nischengruppen. Daher kann auch die Werbung, die im Zusammenhang mit diesen Nischeninhalten gezeigt wird, an die kleinere Zielgruppe besser angepasst werden. Auch kleinere Anbieter können aufgrund geringerer Kosten hier Werbung schalten. Weiterhin erlaubt die Übertragung über IP die Lokalisation der einzelnen Empfänger, sodass hier auch Werbeinhalte lokal angepasst werden. Für Übertragungen per Unicast ist es theoretisch möglich, für jeden Empfänger individuelle Werbung anzubieten. Auch kann der Rückkanal im IPTV dazu genutzt werden, dem Konsumenten die Möglichkeit zu geben, Feedback über die Werbung zu geben und so deren Inhalte noch exakter auf ihn abzustimmen [Chr05].

Allerdings ist die personalisierte Werbung per IPTV nicht ganz unproblematisch. Bei einer Unicast-Übertragung kann genau festgestellt werden, welche Programme ein Benutzer konsumiert. Auch kann ermittelt werden, ob und bei welcher Werbung er möglicherweise das Programm wechselt. So kann ein genaues Profil des Kunden erzeugt werden. Hier müssen entsprechende Datenschutzmaßnahmen ergriffen werden, sofern der Kunde nicht mit der Nutzung solcher Daten einverstanden ist [Chr05].

6 Zusammenfassung und Ausblick

Zusammenfassend lässt sich sagen, dass IPTV mehr ist, als nur ein zusätzlicher Verbreitungskanal neben Kabel, Satellit und terrestrischer Übertragung. Das Internet als frei zugänglicher Übertragungsweg ermöglicht eine Reihe neuer Geschäftsmodelle, die sich deutlich vom klassischen Fernsehen unterscheiden, wie beispielsweise Joost. Dienste wie Video-on-Demand könnten das Fernsehverhalten grundlegend verändern. Auch ist durch IPTV eine größere An-

zahl von Kanälen möglich, sodass es mehr Nischenprogramme gibt. Für die Werbung ergeben sich durch Lokalisierung und Personalisierung der Inhalte ebenfalls neue Möglichkeiten.

Für die Zukunft wird IPTV ein großes Potenzial vorhergesagt. Studien gehen für das Jahr 2010 von 36 bis 63 Mio. Haushalten und einem Marktvolumen zwischen 12 und 27 Mrd. $ aus [CoLo06, S. 6], wobei diese Zahlen die Übertragung über das Internet noch nicht beinhalten. Treiber ist hier die zunehmende Versorgung privater Haushalte mit Breitbandinternetzugängen. Bisher sieht es, zumindest in Deutschland, noch anders aus. Bis Ende 2007 peilt T-Home 100.000 bis 200.000 Kunden an [Dig07].

Interessant wird sein, wie IPTV sich auf Rundfunknetzbetreiber und TV-Sender auswirkt. Erstere bekommen Konkurrenz durch Telekommunikations- und Internetanbieter, die IPTV neben Telefon und Internet über ihre Netze anbieten können. Man spricht hier auch von Triple Play. Auf der anderen Seite bieten Kabelnetztreiber zunehmend Internet über ihre Leitungen an. TV-Sender dagegen bekommen Konkurrenz von zahlreichen Spartensendern, die mit IPTV empfangbar sind und mit denen sie um Werbeeinnahmen konkurrieren müssen. Auf der anderen Seite erhalten die TV-Sender neue Möglichkeiten, ihre Inhalte zu verwerten. So hat IPTV das Potenzial, den heutigen Fernsehmarkt grundlegend zu verändern.

Literaturverzeichnis

[And04] Anderson, C.: The Long Tail, Wired 2004. http://www.wired.com/wired/archive/12.10/tail.html. Abruf 2007-08-28.

[Cat07a] Catone, J.: Internet Killed The Television Star: Reviews of Josst, Babelgum, Zatto and More. Read/WriteWeb 2007. http://www.readwriteweb.com/archives/internet_killed_the_television_star_joost_babelgum_zattoo.php. Abruf 2007-08-23.

[Cat07b] Catone, J.: Is The TV Channel Dead?. Read/Write/Web 2007. http://www.readwriteweb.com/archives/is_the_tv_channel_dead.php. Abruf 2007-08-27.

[Chr05] Christian, P.: Addressable IPTV Advertising: Dramatically Increasing Advertising Renvenue per Viewer. IP Television Magazine 2005. http://www.iptvarticles.com/iptvmagzine_2005_10_addressable_iptv_advertising.htm. Abruf 2007-30-08.

[CoLo06] Cooper, W.; Lovelace, G.: IPTV-Guide- Delivering audio and video over broadband. Informativ 2006. http://informitv.com/reports/iptv/guide/. Abruf 2007-08-24.

[Dig07] Digital fernsehen. IPTV bleibt weiterhin Telekom-Sorgenkind. Digital fernsehen 2007. http://www.digitalfernsehen.de/news/news_186381.html. Abruf 2007-09-01.

[Goo05] Good, Robin: IPTV vs. Internet TV: Key Differences. Robin Good 2005. http://www.masternewmedia.org/2005/06/04/iptv_vs_internet_television_key.htm. Abruf 2007-08-28.

[Gug06] Gugel, Bertram: IPTV vs. InternetTV. Gugelproductions 2006. http://www.gugelproductions.de/blog/2006/iptv-vs-internet-tv.html. Abruf 2007-08-24.

[ITU07] International Telecommunication Union. IPTV Vocabulary of Terms. FG IPTV-DOC-0132. ITU 2007.

[Joo] Joost: http://www.joost.com. Abruf 2007-08-29.

[Kar06] Karsten, E.: Fernsehen Digital – Eine Einführung. Wiesbaden 2006.

[Mac07] MacManus, R.: TVNZ ondemand Being Positioned as „Third Channel“. Read/WriteWeb 2007. http://www.readwriteweb.com/archives/tvnz_ondemand.php. Abruf 2007-08-23

[Tho] T-Home: http://www.t-home.de. Abruf 2007-31-08.

[Tvn] TVNZ ondemand: http://tvnzondemand.co.nz. Abruf 2007-08-29.

[Wik] Wikipedia: IPTV. http://de.wikipedia.org/wiki/IPTV.

Die Infrastruktur von Suchmaschinen am Fallbeispiel Google

Philipp Ciechanowicz

1 Einleitung

Einen Großteil der Zeit, die ein Benutzer im Internet verbringt, verwendet er darauf, etwas zu suchen: Seien es Bilder, Videos oder Webseiten, der Benutzer ist häufig dazu angehalten, aus einem Fundus an Informationen die für ihn passende zu finden. In den frühen Jahren des World Wide Webs haben Verzeichnisse bzw. Portale die wachsende Anzahl an verfügbaren Informationen versucht zu kategorisieren, um dem Benutzer den Einstieg ins WWW zu erleichtern [15]. Mit der rasant wachsenden Anzahl an Webseiten scheint dieser Versuch jedoch aussichtslos, da Portalbetreiber unmöglich sämtliche Seiten manuell indexieren und verlinken können. Dieser Umstand hat hauptsächlich dazu beigetragen, dass Suchmaschinen ihren Einzug ins Web gehalten haben. Diese indexieren automatisiert sämtliche im Internet verfügbaren Dokumente und machen sie damit durchsuchbar. Eine Suche liefert i. d. R. innerhalb von Sekundenbruchteilen Ergebnisse zurück – und das weltweit auf einer 24/7/365-Basis.

Der vorliegende Beitrag beleuchtet die von Suchmaschinenbetreibern eingesetzte Infrastruktur, die zur Bereitstellung dieses hochperformanten Dienstes benötigt wird. Als Fallbeispiel wurde der Suchmaschinenbetreiber Google gewählt, da dieser zum einen Marktführer [14] und zum anderen hinreichend dokumentiert ist [1]. Die Ausführungen sind jedoch so allgemein wie möglich gehalten, sodass an diversen Stellen auch auf Konkurrenten wie Live Search von Microsoft und Yahoo eingegangen wird. Während sich die Abschnitte 2.1 und 2.2 mit der eingesetzten Hard- und Software technischen und ökonomischen Aspekten widmen, diskutiert Kapitel 3 die von Google angebotenen Dienstleistungen unter sozialen Gesichtspunkten.

2 Google's Infrastruktur

Google's Infrastruktur ist in vielerlei Hinsicht extrem: Neben einer enormen Anzahl an Servern bietet Google einen weltweit verfügbaren Dienst an, der Ergebnisse innerhalb von Sekundenbruchteilen liefert. Google's Erfolg ist dabei kein Zufall, sondern begründet sich auf die folgenden Prinzipien [4]:

- Der Einsatz von Standardhardware ist günstiger als der Einsatz von High-End-Hardware, wodurch die Kosten pro Berechnung reduziert werden. Durch die weltweite 24/7/365-Verfügbarkeit ist die Suchmaschine auf hohe Durchsatzraten optimiert, sodass das Verhältnis von Preis/Rechenleistung wichtiger ist als die Maximalleistung (vgl. Abschnitt 2.1.1).
- Die geforderte Hochverfügbarkeit und Performanz wird durch Replikation statt durch teure Spezialhardware erzielt. Dies spart zum einen Kosten, zum anderen müssten selbst beim

Einsatz von RAID-Systemen und USVs[1] Maßnahmen ergriffen werden, um Netzwerk- oder Festplattenausfälle zu kompensieren (siehe Abschnitt 2.1.2).

- Durch die hohe Anzahl an betriebenen Servern sind Hardwareausfälle eher die Regel als die Ausnahme. Um dennoch einen hohen Grad an Zuverlässigkeit zu gewährleisten, setzt Google auf Softwareebene auf fehlertolerante Anwendungen, die trotz Hardwareausfällen die gewünschten Resultate liefern (siehe Abschnitte 2.2.1 und 2.2.2).

Die folgenden Abschnitte geben einen kurzen Überblick über die von Google eingesetzt Infrastruktur. Diese umfasst zum einen Hardwarekomponenten, die in Abschnitt 2.1 vorgestellt werden, zum anderen Softwarekomponenten, die in Abschnitt 2.2 diskutiert werden.

2.1 Hardware

Im April 2007 hat Google mit einem weltweiten Marktanteil von mehr als 55 % nahezu 3,8 Mrd. Suchanfragen bearbeitet [14]. Pro Sekunde bearbeitet Google folglich rund 1.500 Suchanfragen und liefert die Ergebnisse i. d. R. innerhalb von 0,5 Sekunden zurück. Um die dafür benötigte Rechenleistung zur Verfügung zu stellen sowie die angestrebte 24/7/365-Verfügbarkeit weltweit zu gewährleisten, betreibt Google nach Schätzungen vom Juni 2006 etwa 60 Rechenzentren[2] mit rund 450.000 Servern [6, 13]. Die Rechenzentren sind über den gesamten Globus verteilt, um eine erste Lastverteilung der Anfragen vorzunehmen. Der Aufbau sowie die Funktionsweise eines Rechenzentrums wird in Abschnitt 2.1.1 vorgestellt, das Design eines Serverracks in Abschnitt 2.1.2. Neben diesen technischen Aspekten werden mit der Frage nach der Stromversorgung in Abschnitt 2.1.3 ökonomische Gesichtspunkte diskutiert.

2.1.1 Aufbau der Rechenzentren

Zur Beantwortung einer Suchanfrage werden im Wesentlichen fünf verschiedene Servertypen eingesetzt (vgl. Abbildung 1):

- Webserver stellen die Schnittstelle zwischen dem Benutzer und dem Rechenzentrum dar. Sie nehmen Suchanfragen entgegen, koordinieren deren Ausführung und liefern das Ergebnis als HTML-Seite zurück.
- Indexserver speichern den invertierten Index und bestimmen zu einer Liste aus Stichwörtern sämtliche Dokumente, die diese Stichwörter enthalten.
- Dokumentenserver speichern den Inhalt aller indexierten Webseiten und bestimmen zu einer Liste von Dokumenten die Überschriften und Textauszüge, die auf der Ergebnisseite angezeigt werden.
- Parallel zur Suche übernimmt ein weiterer Server die Rechtschreibprüfung und unterbreitet evtl. Verbesserungsvorschläge („Meinten Sie:“).

---

1 Unabhängige Stromversorgung.

2 Aktuell baut Google unter dem Codenamen *Project 02* in The Dalles, Oregon ein Rechenzentrum, während Microsoft und Yahoo 130 km nördlich ebenfalls Rechenzentren errichten.

- Parallel zur Suche übernimmt bei Google und Yahoo ein weiterer Server die Generierung von kontextabhängiger Werbung, über die sich beide Suchmaschinenbetreiber hauptsächlich finanzieren. Folglich entfällt dieser Schritt bei Live Search von Microsoft.

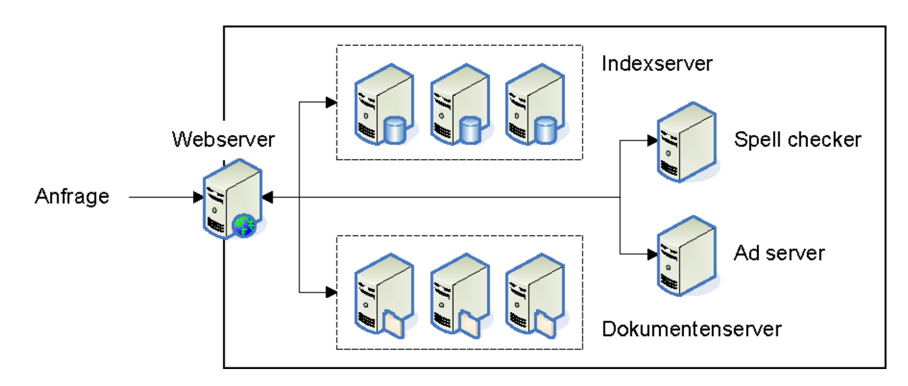


Abbildung 1: Vereinfachter Aufbau eines Rechenzentrums von Google

Die Bearbeitung einer Suchanfrage beginnt schon mit der Eingabe der URL und der damit verbundenen Namensauflösung. Google setzt hier einen DNS-basierten Lastverteilungsmechanismus ein, der sowohl die geografische Lage des Benutzers als auch die Auslastung der Rechenzentren berücksichtigt, um die Round Trip Time (RTT) zu minimieren. Darüber hinaus führt jedes Rechenzentrum eine Lastverteilung zwischen den verfügbaren Webservern durch. Nachdem ein Rechenzentrum und ein Webserver ausgewählt wurden, wird die Bearbeitung der Suchanfrage vollständig lokal durchgeführt. In einem ersten Schritt sendet der Webserver die Anfrage an die Indexserver, die eine Liste von *docIds* ermitteln. Diese Liste identifiziert sämtliche Dokumente, die der Suchanfrage genügen, und kann parallel von mehreren Servern erstellt werden: Zum einen können sämtliche Wörter in der Anfrage trivialerweise parallel bearbeitet werden, zum anderen können aber auch einzelne Wörter parallel bearbeitet werden, da der invertierte Index in mehrere so genannte *Shards* unterteilt ist, die ebenfalls parallel durchsucht werden können. Abschließend wird die Liste gemäß den vom Benutzer eingegebenen Boolschen Operatoren verknüpft und an den Webserver zurückgegeben. In einem zweiten Schritt sendet der Webserver die docIds an die Dokumentenserver. Diese generieren zu jedem Dokument die Überschrift, den Textauszug sowie die URL, die auf der Ergebnisseite angezeigt werden. Auch dieser Schritt ist trivialerweise parallelisierbar, da die Überschrift, der Textauszug und die URL von verschiedenen Servern erstellt werden können. Parallel zu diesen beiden Phasen wird die Rechtschreibung überprüft und kontextspezifische Werbung[3] generiert. Der finale Bearbeitungsschritt wird vom Webserver durchgeführt, indem er sämtliche Ergebnisse auf eine HTML-Seite zusammenfasst und diese an den Aufrufer zurückgibt.

2.1.2 Aufbau der Serverracks

Google setzt in seinen Rechenzentren relativ günstige Server ein, die rund 1.000 $ pro Stück kosten und nach Spezifikationen von Google gefertigt werden. Aktuell sind von 533 MHz Intel Celeron bis zu dual 1,4 GHz Intel Pentium III mehrere CPU-Generationen im Einsatz, wobei Google aufgrund des relativ hohen Stromverbrauchs der Intel-Prozessoren mittlerweile auch Opteron-Prozessoren von AMD einsetzt (siehe Abschnitt 2.1.3). Die Server sind jeweils mit

---

[3] Zumindest bei Google und Yahoo.

2 GB RAM sowie einer oder mehrerer 80 GB-Festplatten bestückt. Auf allen Servern läuft ein modifiziertes Red Hat Linux-Betriebssystem, wobei Google teilweise eigene Patches entwickelt hat. Yahoo benutzt hauptsächlich Free BSD, Linux und Windows, Microsoft natürlich ausschließlich hauseigene Produkte.

Die Server werden in den Rechenzentren in so genannten *Racks* zusammengeschlossen. Ein Rack bietet dabei auf jeder Seite Platz für 20 2u- bzw. 40 1u-Server, sodass ein Rack insgesamt 40 bis 80 Server aufnehmen kann.[4] Die Server einer Rackseite sind per 100 Mbps-Ethernet an einen Switch gekoppelt, wobei jeder Switch per 2 Gbps-Ethernet an einen Core-Switch gekoppelt ist, der sämtliche Racks eines Rechenzentrums miteinander verbindet.

	Standardserver	High-End-Server
CPU	88 · dual 2 GHz Intel Xeon	8 · 2 GHz Intel Xeon
RAM	jeweils 2 GB = 176 TB	64 GB
HDD	jeweils 80 GB ≈ 7 TB	8 TB
Preis	278.000 $	758.000 $

Abbildung 2: Preisvergleich zwischen Standard- und High-End-Server

Im Gegensatz zu Microsoft, bei denen wenige teure Hochleistungsserver zum Einsatz kommen, setzt Google viele billige Standardserver ein. Ein Preisvergleich aus dem Jahr 2002 [4] verdeutlicht Google's Prinzip: Der Hochleistungsserver ist fast dreimal so teuer, hat dreimal weniger Hauptspeicher, deutlich weniger CPUs und geringfügig mehr Festplattenkapazität als der Standardserver (vgl. Abbildung 2). Der Preis ist zwar durch die höhere Qualität der Komponenten und die schnellere Interkonnektivität zwischen den CPUs gerechtfertigt. Da Google's Architektur aber auf Redundanz setzt, sind diese beiden Eigenschaften den höheren Anschaffungspreis nicht wert.

2.1.3 Stromversorgung

Der Stromverbrauch spielt bei Google's Rechenleistung eine große Rolle. Bis ins Jahr 1999 hat Google seine Server noch in professionellen Rechenzentren betreiben lassen, bei den aktuellen Leistungsdichten der von Google eingesetzten Server stoßen Rechenzentren jedoch an ihre Leistungsgrenzen. Das folgende Rechenbeispiel verdeutlichet dies [4]: Ein dual 1,4 GHz Intel Pentium III-Prozessor benötigt unter Volllast 90 W (55 W für die beiden CPUs, 10 W für die Festplatte und 25 *W* für den Hauptspeicher und das Mainboard). Bei einer Energieeffizienz von 75 %, wie sie für ein ATX-Netzteil üblich ist, benötigt ein Server folglich 120 W, sodass ein Rack mit 80 Servern etwa 10 kW benötigt. Bei einer Rack-Grundfläche von 25 ft^2 ergibt sich somit eine Leistungsdichte von 400 W/ft^2, die bei leistungsfähigeren Prozessoren auch schnell bis zu 700 W/ft^2 betragen kann.[5] Problematisch ist die Tatsache, dass typische Rechenzentren Leistungsdichten zwischen 70 W/ft^2 und 200 W/ft^2 tolerieren, sodass Google gezwungen war, eigene Rechenzentren zu betreiben.

---

4 1u = 1,75 Zoll = 4,45 cm ist die standardisierte Höheneinheit für Server.

5 1 ft = 30,48 cm.

Mit der veranschlagten Leistungsaufnahme von 10 kW pro Rack verbraucht ein Rack im Monat 10 kW · 24 h · 30 d ≈ 10 MWh Strom. Bei einem in den USA großzügig kalkulierten Preis von 15 Cent/kWh belaufen sich die monatlichen Stromkosten folglich auf 1.500 $, sodass es vordergründig sehr erstrebenswert scheint, auf jeden Fall sparsamere Prozessoren einzusetzen. Diese sind jedoch i. d. R. unverhältnismäßig teuer, sodass eine Reduzierung des Stromverbrauchs per se unwirtschaftlich wäre, da die monatliche Abschreibung eines Serverracks bei einem Abschreibungszeitraum von drei Jahren mit 7.700 $ das Fünffache beträgt (vgl. Abschnitt 2.1.2). Als Entscheidungskriterium setzt Google deshalb die Leistung des Prozessors in Relation zu dessen Stromverbrauch. Diese Kennzahl ist bei den aktuellen Prozessoren von AMD offensichtlich größer, weshalb Google seit neuestem Opteron-Prozessoren in seinen Servern einsetzt.

2.2 Software

Mit seinen geschätzt mehr als 450.000 Servern hat Google eine enorme Infrastruktur aufgebaut, die naturgemäß anfällig für Hardwareausfälle ist. Selbst bei einer optimistisch geschätzten durchschnittlichen Ausfallrate von einem Ausfall innerhalb von drei Jahren würden bei Google täglich mehr als 410 Server ausfallen. Hardwaredefekte sind folglich keine Ausnahme, sondern die Regel. Dieser Unzuverlässigkeit auf Hardwareebene begegnet Google auf der Softwareebene mit fehlertoleranten Anwendungen, die trotz Serverabstürzen, Netzwerk- und Stromausfällen die gewünschten Resultate liefern.

In den folgenden Abschnitten werden exemplarisch zwei Anwendungen vorgestellt, die bei Google derzeit produktiv im Einsatz sind: Abschnitt 2.2.1 befasst sich mit dem von Google entwickelten Dateisystem, während Abschnitt 2.2.2 eine Anwendung zur Programmierung von Parallelrechnern vorstellt.

2.2.1 Google File System

Das Google File System (GFS) ist ein im Jahr 2002 entwickeltes auf Linux basierendes verteiltes Dateisystem, das durch Replikations- und Lastverteilungsmechanismen eine hohe Verfügbarkeit, Ausfallsicherheit und Zuverlässigkeit aufweist [10]. Ursprünglich aus dem *BigFiles*-Projekt hervorgegangen, wird GFS mittlerweile auf sämtlichen Produktivservern von Google eingesetzt.

Ein GFS-Cluster besteht immer aus genau einem *Masterserver* sowie mehreren *Chunkservern*. Die Chunkserver speichern die eigentlichen Nutzdaten, die in Blöcke von 64 MB aufgeteilt werden, den so genannten *Chunks*. Jedes dieser Chunks ist global durch einen 64-stelligen Bitstring eindeutig identifizierbar, der beim Erzeugen des Chunks vom Masterserver vergeben wird. Um die geforderte Hochverfügbarkeit zu gewährleisten, existieren von sämtlichen Chunks mindestens drei Repliken, bei häufig zugegriffenen Dateien teilweise deutlich mehr. Der Masterserver hingegen speichert lediglich Metadaten über die Chunks, wie z. B. Namensräume, das Mapping von Dateien zu Chunks oder den Speicherort der Repliken. Er ist darüber hinaus für die Erstellung von Repliken sowie für die Handhabung von Hardwarefehlern zuständig.

Clients müssen sich beim Anfordern einer Datei zunächst an den Masterserver wenden. Dieser ermittelt einen Chunkserver, der die gewünschte Datei vorhält, und übermittelt dem Client dessen Adresse. Intern führt der Master dabei eine Lastverteilung durch, da er die Auslastung der einzelnen Chunkserver berücksichtigt. Die anschließende Kommunikation findet ausschließlich zwischen dem Client und dem Chunkserver statt, sodass der Masterserver keinen Engpass darstellt. Um den Masterserver weiter zu entlasten und die Zugriffsgeschwindigkeit zu erhöhen, cachen Clients Adressmetadaten über die Chunkserver, sodass zukünftige Zugriffe evtl. ohne den Masterserver stattfinden können.

Die Gewährleistung der Ausfallsicherheit und Hochverfügbarkeit ist grund-sätzlich Aufgabe des Masterservers. Durch so genannte *Herzschlagnachrichten* (engl. heart beat messages) überprüft der Masterserver periodisch die Verfügbarkeit von Chunkservern. Ist ein Chunkserver ausgefallen, leitet der Masterserver einen Replikationsprozess ein. Dabei verfolgt er folgende Zielsetzungen:

- Repliken sollen auf Chunkservern erstellt werden, die eine unterdurchschnittliche Festplattenauslastung aufweisen.
- Chunks müssen über mindestens drei Serverracks verteilt werden, um Stromausfällen und Netzwerkproblemen vorzubeugen.
- Die Anzahl an „kürzlich“ erzeugten Repliken auf einem Chunkserver ist so gering wie möglich zu halten.

Die Gewährleistung der Datenintegrität ist Aufgabe der Chunkserver. Bevor diese Ergebnisse an Clients zurückgeben, wird eine Prüfsumme berechnet. Falls diese nicht korrekt ist, wird eine Fehlernachricht an den Client zurückgegeben und der Masterserver von der fehlerhaften Replik in Kenntnis gesetzt. Der Client liest die Datei folglich von einem anderen Chunkserver, während der Masterserver „innerhalb von Minuten“ das Erstellen einer neuen Replik einleitet.

2.2.2 MapReduce

MapReduce bezeichnet ein von Google im Frühjahr 2003 entwickeltes Programmiermodell sowie dessen konkrete Implementierung als C++-Bibliothek [9]. Ziel der Bibliothek ist die automatisierte Parallelisierung von Berechnungen, die auf enormen Datenmengen operieren. Die Details der Parallelisierung werden dabei von der Bibliothek gekapselt, sodass selbst mit der Programmierung von parallelen Anwendungen unerfahrene Benutzer die enorme Rechenleistung des Google-Clusters nutzen können. Während die Bibliothek die Verteilung der Daten vornimmt, die Ausführung auf verschiedenen Prozessoren koordiniert, Hardwarefehler sowie die vollständige Kommunikation handhabt, muss der Benutzer lediglich zwei Funktionen *map* und *reduce* spezifizieren. Auf diese Weise lässt sich eine Vielzahl von Aufgaben effizient lösen.

Das zugrunde liegende Konzept von MapReduce ist vergleichsweise simpel. Auf eine Menge von *x* Datensätzen wird sukzessive eine map-Funktion angewendet, die als Argument ein Schlüssel-Wert-Paar erwartet und aus diesem eine Liste aus Schlüssel-Wert-Paaren erzeugt: *map(k1,v1)→Liste(k2,v2)*. An der Signatur ist ersichtlich, dass sowohl die Schlüssel als auch die Werte der Eingabeparameter nicht mit denen der Ausgabe übereinstimmen müssen. Bevor das Zwischenergebnis an die reduce-Funktion weitergegeben wird, gruppiert die Bibliothek sämtli-

che Schlüssel-Wert-Paare, die denselben Schlüssel aufweisen, und sortiert diese in lexikografischer Reihenfolge. Auf diese Weise entsteht eine Liste, bei der jeder Schlüssel einmalig ist und dem eine Menge von Werten zugeordnet ist: *group(Liste(k2,v2))→(k2,Liste(v2))*. Die reduce-Funktion erwartet als Parameter eben diese Liste aus Schlüssel-Werte-Paaren und transformiert diese in eine Liste aus Werten: *reduce(k2,Liste(v2))→Liste(v3)*. Diese Liste aus Werten stellt das Endergebnis der MapReduce-Operation dar.

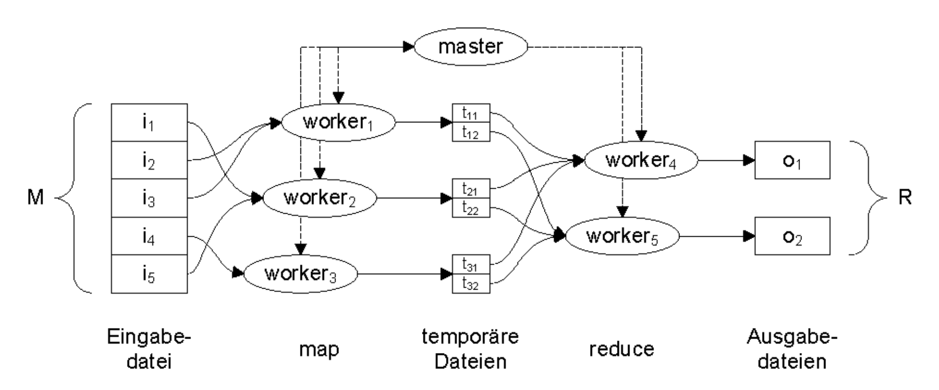


Abbildung 3: Ablauf einer MapReduce-Operation mit M=5 und R=2

Aufgabenstellungen, die sich mit dem oben beschriebenen Konzept lösen lassen, sind trivialerweise parallelisierbar, da sowohl das Ergebnis der map- als auch der reduce-Funktion unabhängig von vorherigen Ergebnissen ist. MapReduce parallelisiert solche Aufgabenstellungen, indem die Arbeit auf mehrere Workerprozesse aufgeteilt wird, die von einem dedizierten Masterprozess koordiniert werden (vgl. Abbildung 3). Bei einer MapReduce-Operation wird zunächst die Eingabedatei in M Teile i_1 – i_M aufspaltet, als Endergebnis werden R Ausgabedateien o_1 – o_R erzeugt. Folglich müssen insgesamt M map-Funktionen und R reduce-Funktionen aufgerufen werden. In einer ersten Phase weist der Masterprozess jeden Teil der Eingabedatei genau einem map-Worker zu, wobei ein map-Worker durchaus mehrere Teile zugewiesen bekommen kann. Jeder map-Worker wendet auf sämtliche ihm zugewiesenen Datensätzen die map-Funktion an und speichert die Zwischenergebnisse lokal in einer temporären Datei. Diese Datei wird, da sie u. U. sehr groß werden kann, mithilfe einer Partitionierungsfunktion in R Teile aufgespaltet. Sobald ein map-Worker sämtlich ihm zugewiesenen Datensätze bearbeitet hat, teilt dieser dem Masterprozess den Speicherort der temporären Dateien mit. Der Masterprozess verteilt diese, analog zur ersten Phase, unter den verfügbaren reduce-Workern, wobei deren Anzahl über den Parameter R gesteuert wird. In einer zweiten Phase wendet jeder reduce-Worker die reduce-Funktion auf sämtliche ihm zugewiesenen Datensätze an. Da jeder reduce-Worker sein Ergebnis in genau einer Ausgabedatei speichert, werden als Endergebnis R Ausgabedateien erzeugt. Diese werden i. d. R. nicht kombiniert, da sie häufig als Eingabe für weitere MapReduce-Operationen dienen.

Getreu dem Designprinzip, Zuverlässigkeit auf der Softwareebene bereitzustellen, beinhaltet MapReduce diverse Mechanismen, um mit Hardwareausfällen umzugehen. Zu diesem Zweck pingt der Masterprozess sämtliche Workerprozesse periodisch an. Erhält er innerhalb einer Zeitspanne keine Antwort, wird der Prozess aus der Liste der verfügbaren Prozesse gestrichen. Falls ein map-Worker betroffen ist, müssen sämtliche Datensätze, die diesem Worker zugewiesen wurden, erneut bearbeitet werden, da etwaige Zwischenergebnisse lediglich lokal bei dem ausgefallen map-Worker vorhanden sind. Falls ein reduce-Workers betroffen ist, müssen lediglich die noch zu bearbeitenden temporären Dateien erneut zugewiesen werden, da etwaige Zwi-

schenergebnisse global im GFS gespeichert werden (vgl. Abschnitt 2.2.1). Falls der Masterprozess von einem Ausfall betroffen ist, wird die MapReduce-Operation abgebrochen. Theoretisch könnte der Zustand des Masterprozesses periodisch gespeichert werden und im Fall eines Absturzes wiederhergestellt werden. Da jedoch lediglich ein Masterprozess zum Einsatz kommt, ist die Wahrscheinlichkeit eines Ausfalls im Vergleich zum Ausfall eines Workerprozesses sehr gering.

MapReduce findet bei Google in diversen Produkten Anwendung. An erster Stelle sei hier die Erstellung des invertierten Index genannt, der Google's Suchalgorithmus zugrunde liegt. Darüber hinaus lässt sich mit MapReduce eine Vielzahl von Aufgabenstellungen bearbeiten, wie z. B. das Zählen von Wörtern in einer Sammlung von Dokumenten. Microsoft's Pendant zu MapReduce nennt sich *Dryad* und verfolgt im Wesentlichen dieselbe Zielsetzung [12]. Im Gegensatz zu Google's Ansatz sieht Dryad jedoch eine nahtlose Integration mit den SQL-Server Integration Services (SSIS) sowie eine deutlich flexiblere Kommunikationstopologie vor. Auf diese Weise lassen sich die Workerprozesse beispielsweise ring- oder sternförmig anordnen.

3 Datenschutz bei Google

„We cannot even answer the most basic questions because we don't know enough about you", bekundete Google CEO Eric Schmidt in einem Interview mit der Financial Times [8]. Die in London ansässige Bürgerrechtsorganisation Privacy International sieht dies jedoch völlig anders. In ihrem Bericht von Juni 2007 [11], in dem sie den Umgang mit personenbezogenen Daten von 23 Internetdienstleistern bewertet hat, stellt Google das Schlusslicht dar: Während Microsoft mit der Note 4 (serious lapses) und Yahoo mit einer 5 (substantial threat) bewertet wurden, erhielt Google eine 6 (hostile to privacy).

In diese Situation hat sich Google durch eigenes Verschulden manövriert. Sei es G-Mail, das den Inhalt sämtlicher E-Mails analysiert, die von einem Konto gesendet oder empfangen werden, Google's Internetsuche, die sämtliche Suchanfragen mithilfe eines Cookies einem Browser eindeutig zuordnen kann, oder das soziale Netzwerk Orkut, das naturgemäß sehr persönliche Daten benötigt, kein anderer Internetdienstleister legt einen derartigen Elan bei der Sammlung von Daten an den Tag. Was Google mit diesen Informationen anfängt, ist vage in einer Datenschutzerklärung festgehalten. Hier heißt es unter anderem: „We may combine the information you submit under your account with information from other Google services or third parties [...]" [2]. Unter Berücksichtigung der Dienste, für die derzeit ein Google-Account benötigt wird[6], ist Google damit faktisch in der Lage, problemlos einen erheblichen Teil des Privatlebens eines Benutzers zu rekonstruieren.

Dass Google auf Belange der informationellen Selbstbestimmung nicht allzu großen Wert legt, verdeutlicht der in Google Maps integrierte Dienst Google Street View, der Bilder von Straßenansichten einiger amerikanischer Großstädte zeigt: Die Qualität der Aufnahmen ist teilweise so gut, dass Personen und Nummernschilder erkennbar sind. Auch das Urheberrecht legt Google bei seinem Buchdigitalisierungsprojekt Google Book Search zu seinen Gunsten aus: Es ge-

---

6 Dazu zählen Blogger, Google Alerts, Google Calendar, Google Mail, Google News Alerts, Google Reader, Google Sitemap, Orkut und die Shoppingliste von Froogle.

währt Autoren und Verlegern das Recht, ihre Werke per Opt-out-Verfahren von der Digitalisierung auszuschließen. Da das Urheberrecht in diesem Fall aber genau den umgekehrten Weg vorschreibt, nämlich eine explizite Erlaubnis des Rechteinhabers, stellt das von Google praktizierte Vorgehen in den Augen vieler Juristen eine illegale Rechtsverdrehung dar.

Unter Berücksichtigung der bereits gesammelten Daten stellt sich die Frage, wie viele Daten Google über seine Benutzer noch sammeln will. Die Antwort gibt Eric Schmidt: „The goal is to enable Google users to be able to ask the questions such as ‚What shall I do tomorrow?' and 'What job shall I take?'" [8]. Dass dafür enorme Mengen an persönlichen Informationen notwendig sind, steht zweifelsohne außer Frage.

4 Zusammenfassung

Die vorliegende Ausarbeitung hat sich mit der Infrastruktur von Suchmaschinen beschäftigt und diese am Fallbeispiel des Marktführers Google diskutiert. Das Unternehmen betreibt seine geschätzt 450.000 Server mittlerweile in eigenen Rechenzentren, da übliche Rechenzentren nicht auf die von Google geforderten Leistungsdichten ausgelegt sind (siehe Abschnitt 2.1.3). Neben dem Aufbau und der Funktionsweise eines von Google betriebenen Rechenzentrums wurde auch die verwendete Hardware vorgestellt (siehe Abschnitte 2.1.1 und 2.1.2). Die besteht zumeist aus billigen Standardkomponenten und ist dementsprechend fehleranfällig, sodass Hardwareausfälle eher die Regel als die Ausnahme sind. Um trotzdem die geforderte Hochverfügbarkeit und Performanz zu gewährleisten, setzt Google auf Replikation statt auf teure Spezialhardware, wie z. B. RAID-Systeme oder USVs. Der so erzielte Kostenvorteil im Vergleich zu Microsoft wird von Experten auf ein Verhältnis von 3:1 geschätzt, d. h., für jeden 1 $, den Google ausgibt, muss Microsoft 3 $ ausgeben, um Schritt zu halten [6].

Neben der Hardware wurde besonderes Augenmerk auf die von Google verwendete Software gelegt. Google verfolgt hier das Credo, Zuverlässigkeit auf Softwareebene zu gewährleisten und fehlertolerante Software zu schreiben, die in einer Umgebung von Serverabstürzen, Netzwerk-, Festplatten- und Stromausfällen trotzdem die gewünschten Ergebnisse liefert. Neben dem Google File System, das im Wesentlichen die Replikation von Dateien vornimmt und damit dessen Verfügbarkeit sicherstellt, wurde mit MapReduce eine Bibliothek vorgestellt, die das Programmieren von Parallelrechnern wesentlich vereinfacht (siehe Abschnitte 2.2.1 und 2.2.2). Darüber hinaus existiert noch eine Reihe von weiteren Anwendungen, die Google vor dem Hintergrund seiner speziellen Anforderungen entwickelt hat, wie z. B. die verteilte Datenbank *BigTable* [7] oder *Chubby*, ein Sperrmechanismus für verteilte Dateien [5].

Für die Zukunft ist davon auszugehen, dass alle Suchmaschinenbetreiber ihre Serverkapazitäten drastisch ausbauen werden, um neue Dienste in hoher Qualität anbieten zu können. Für einen Überblick über die von Google derzeit angebotenen Dienste sei an dieser Stelle auf [3] verwiesen. Während Google hier auf Standardhardware und Replikation setzt, geht Microsoft mit teurer Spezialhardware einen anderen Weg. Welches der beiden Konzepte sich schlussendlich durchsetzt, ist aus Benutzersicht erst einmal irrelevant, da für diesen in erster Linie die Qualität der Suchergebnisse zählt. Wie sich der Benutzer diese Ergebnisse allerdings *erkauft*, steht auf einem anderen Blatt. Vorsicht ist hier vor allem bei der von Google exzessiv betriebe-

nen Sammlung von Benutzerdaten geboten, da diese durch das Verknüpfen der Datenpools faktisch zur Rekonstruktion von Persönlichkeitsprofilen herangezogen werden kann.

Literaturverzeichnis

[1] Google Labs Research Publications. http://labs.google.com/papers.html.

[2] Google privacy policy. http://www.google.com/privacypolicy.html.

[3] Thor Alexander. Guck bei Google! Das Google-Lexikon. Rhombos, 2007.

[4] Luiz André Barroso, Jeffrey Dean, and Urs Hölzle. Web Search for a Planet: The Goolge Cluster Architecture. IEEE Micro, 23(2):22-28, 2003.

[5] Mike Burrows. The Chubby lock service for loosely-coupled distributed systems. In 7th Symposium on Operating Systems Design and Implementation, 2006.

[6] David F. Carr. How Google Works. Baseline Magazine, 06. Juli 2006. http://www.baselinemag.com/article2/0,1540,1985040,00.asp.

[7] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, and Robert E. Gruber. Bigtable: A Distributed Storage System for Structured Data. In 7th Symposium on Operating Systems Desing and Implementation, 2006.

[8] Caroline Daniel and Maija Palmer. Google's goal: to organise your daily life, 22. Mai 2007. http://www.ft.com/cms/s/2/c3e49548-088e-11dc-b11e-000b5df10621,dwp_uuid=e8477cc4-c820-11db-b0dc-000b5df10621.html.

[9] Jeffrey Dean and Sanjay Ghemawat. MapReduce: Simpified Data processing on Large Clusters. In 6th Symposium on Operting System Desing and Implementation, 2004.

[10] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. The Google File System. In 19th Symposium on Operating Systems Principles, 2003.

[11] Privacy International. A race to the bottom: Privacy ranking of internet service companies, 09. Juni 2007. http://www.privacyinternational.org/issues/internet/interimrankings.pdf.

[12] Michael Isard, Mihai Budiu, Yuan Yu, Andrew Birrell, and Dennis Fetterly. Dryad: Distributed Data-Parallel Programs from Sequential Building Blocks. Technical report, Microsoft Research Silicon Valley, 2006.

[13] John Markoff and Saul Hansell. Hiding in Plain Sight, Google Seeks More Power. The New York Times, 08. Juni 2006. http:// www.nytimes.com/2006/06/14/technology/14search.html? ex=1307937600& en=d96a72b3c5f91c47&ei=5090.

[14] SearchEngineWatch.com. Top 10 Search Providers, April 2007, 19. Juni 2007. http://searchenginewatch.com/showPage.html? page=3626208.

[15] Gottfried Vossen and Stephan Hagemann. Unleashing Web 2.0. Elsevier, 2007.

Amazon-Webservices – Eine Betrachtung technischer, ökonomischer und sozialer Aspekte

Oliver Müller

1 Einleitung

Fragt man Shai Agassi, Vorstandsmitglied von SAP und Leiter des Bereichs Produktentwicklung und Technologie, nach der bedeutendsten Innovation in der Softwareindustrie in den letzten zehn Jahren, so antwortet er nicht Linux, Google oder Web 2.0, sondern: der Amazon-Webservice-Stack.[1]

In Sinne des World Wide Web Consortiums (W3C) ist unter einem Webservice eine Softwarekomponente zu verstehen, die ihre Funktionalität über ein Netzwerk anderen Komponenten oder Anwendungen zur Verfügung stellt. Basierend auf offenen XML-Standards (insb. WSDL, SOAP, UDDI) sowie standardisierten Internetprotokollen (insb. HTTP, TCP/IP) können Webservices in einheitlicherweise beschrieben, veröffentlicht, lokalisiert und aufgerufen werden.[2] Mit dem Paradigma der Serviceorientierung werden u. a. folgende Ziele verfolgt: Aus technischer Sicht besteht die Hoffnung, Entwicklungszeit und -kosten durch die Wiederverwendung von Services in unterschiedlichen Anwendungssystemen einsparen zu können. Zudem soll durch die Komposition lose gekoppelter Services, im Vergleich zu einer monolithischen Architektur, eine erhöhte Flexibilität von Anwendungssystemen erreicht werden. Aus ökonomischer Sicht ermöglicht die Serviceorientierung, Funktionalitäten der eigenen Anwendungssysteme externen Nutzern effizient verfügbar zu machen und so Einnahmen zu generieren.

Amazon, der weltgrößte Onlinehändler, hat elf Jahre und über 2 Mrd. $ in die Entwicklung der eigenen weltweit verteilten IT-Infrastruktur investiert.[3] Im Juli 2002 hat Amazon im Zuge der „Amazon-Webservices-Strategy" begonnen, Teile dieser Infrastruktur zu öffnen und als Services zur Verfügung zu stellen. Jeder, der im Besitz einer Kreditkarte ist, hat seitdem die Möglichkeit, die Ressourcen, Daten und Funktionalitäten der Amazon-Infrastruktur zu nutzen. Als Motivation dieser – für einen Onlinehändler eher unüblichen – Initiative hat Jeff Bezo, CEO von Amazon, auf der Web 2.0-Conference 2006 in San Francisco das Ziel genannt, einen vom Handel unabhängigen Geschäftsbereich aufzubauen, der sich insbesondere an einzelne Softwareentwickler sowie junge, kleine bis mittelständische Unternehmen richtet und letztendlich zusätzliche Einnahmen generieren soll.[4]

---

1 Vgl. Tapscott, Williams (2007).

2 Vgl. Haas, Brown (2004).

3 Vgl. Hof (2006).

4 Vgl. Arrington (2006).

2 Der Amazon-Webservice-Stack

Im Mittelpunkt des Amazon-Webservice (AWS)-Stack steht der Service. Amazon interpretiert den Begriff Webservice recht untypisch und bewusst minimalistisch wie folgt:[5]

Webservice = Business Modell + API

Als Konsequenz aus dieser Definition kapselt ein Amazon-Webservice nicht zwangsweise nur die Funktionalität einer Softwarekomponente, sondern kann sich auch auf Infrastrukturressourcen wie Speicherplatz oder Rechenkapazität sowie menschliche Intelligenz oder Arbeitskraft, wie es bspw. beim Mechanical Turk Service der Fall ist, beziehen. Grundlegend ist lediglich, dass ein angebotener Service einen ökonomischen Nutzen birgt und über ein Application Programming Interface (API), also eine Schnittstelle im Sinne der Softwaretechnik, genutzt werden kann. Zurzeit werden die Amazon-Webservices von über 200.000 Nutzern als Bausteine für die Entwicklung eigner Webseiten und Anwendungen genutzt. Die Bandbreite der Nutzer reicht dabei von einzelnen Softwareentwicklern bis hin zu Softwaregiganten wie Microsoft.[6] Aus Nutzersicht sind die Amazon-Webservices interessant, da sie auf einem niedrigen Preisniveau, welches im Rahmen des Budgets von Start-up-Unternehmen oder Einzelpersonen liegt, ausgereifte Funktionalitäten und Ressourcen bereitstellen. Alle Services setzen auf der Infrastruktur auf, die Amazon auch für den Betrieb der eigenen E-Commerce-Plattform nutzt und besitzen folglich eine extrem hohe Zuverlässigkeit, Verfügbarkeit und Skalierbarkeit. Aus dieser Tatsache ergibt sich auch Amazons Motivation für die Webserviceaktivitäten. Jegliche angebotenen Services basieren auf bereits vorhandenen Ressourcen oder Funktionalitäten. Dadurch entstehen Amazon nur minimale Zusatzkosten, wodurch das Geschäftsfeld bereits mit geringen Einnahmen profitabel werden kann.

Der Amazon-Webservice-Stack besteht derzeit aus zehn Services, die Jeff Barr (Amazon-Webservice-Chefentwickler) in die Schichten *Infrastructure as a Service, Data as a Service*, *Search as a Service* und *People as a Service* unterteilt (vgl. Abbildung 1).[7]

Die Schicht **Infrastructure as a Service** beinhaltet Services, die Teile von Amazons enormer, webbasierter IT-Infrastruktur, der sog. *Amazon Cloud*, auf der auch Amazons eigene E-Commerce-Plattform läuft, externen Nutzern zur Verfügung stellt. Der *Simple Storage Service (S3)* (Preis: 0,15 $ je GB Speicherplatz pro Monat plus 0,18 $ je GB Transfer pro Monat) stellt Entwicklern virtuellen Speicherplatz über das Web zur Verfügung. Analog kann über den *Elastic Compute Cloud Service (EC2)* (Preis: 0,10 $ je CPU pro Stunde) die Rechenkapazität der Cloud genutzt werden. Ergänzt werden die beiden Services durch den *Simple Queue Service (SQS)* (Preis: 0,10 $ pro 1.000 Nachrichten plus ca. 0,15 $ pro TB Datentransfer), der einen zuverlässigen Transfer von Nachrichten zwischen verteilten Anwendungen ermöglicht.

Die Schicht **Data as a Service** umfasst zwei Services: Der E-Commerce-*Service (ECS)* stellt Amazons Produktdaten und E-Commerce-Funktionalitäten bereit. Er erlaubt Entwicklern, Webseitenbetreibern und Händlern, die gleichen Daten und Funktionalitäten für ihre Anwendungen

---

[5] Vgl. Barr (2007).

[6] Microsoft nutzt den Simple Storage Service (S3), um kurzfristig Engpässe bei der Verfügbarkeit von Speicherkapazität zu überbrücken, vgl. Arrington (2006).

[7] Vgl. Barr (2007).

zu nutzen, die auch Amazon für sein eignees E-Commerce-Geschäft nutzt. Neben detaillierten Produkt- und Preisinformationen über alle gelisteten Artikel sind bspw. auch Kundenrezensionen und Produktempfehlungen abrufbar. Zudem können die Funktionen Einkaufswagen und Wunschzettel in eigene Anwendungen integriert werden. Der zweite Service dieser Schicht ist der *Historical Pricing Service* (Preis: 249 $ für bis zu 60.000 Zugriffe pro Monat), welcher den Zugriff auf historische Preis- und Abverkaufsdaten der letzten drei Jahre für die Produktgruppen Bücher, Musik und DVDs erlaubt. Diese Informationen können anderen Händlern eine wertvolle Entscheidungsgrundlage zur Preisbildung oder Prognose von Nachfrage- und Preisentwicklungen bieten.

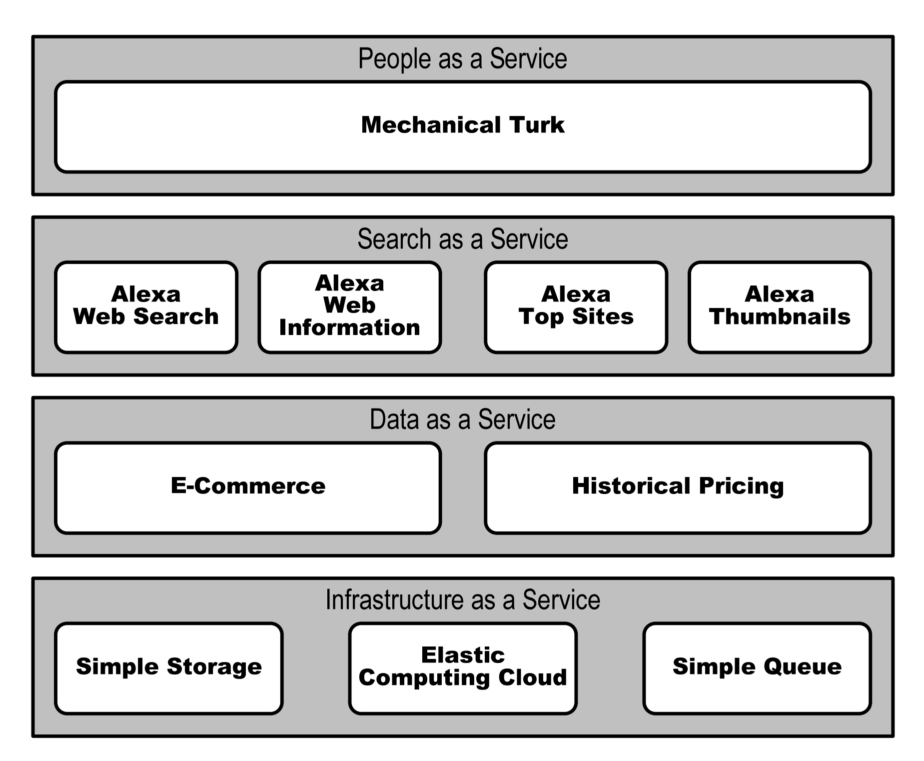


Abbildung 1: Der Amazon-Webservice-Stack[8]

Die Schicht **Search as a Service** umfasst vier Services, die die diversen Funktionalitäten der Alexa Web Search Engine kapseln: Der *Alexa Web Search Service* (Preis: ab 0,30 $ pro 1.000 Anfragen) ermöglicht die Einbindung der Suchfunktion von Alexa in eigene Anwendungen. Der *Alexa Top Sites Service* (Preis: 0,25 $ für 100 URLs) liefert den Rang einer Webseite im sog. Alexa Traffic Ranking, welches aus einer Kombination von Unique Users und Page Views errechnet wird. Der *Alexa Web Information Service* (Preis: 0,15 $ für 1.000 Anfragen) macht Informationen über die Nutzung und Struktur (z. B. historische Nutzungszahlen, Größe einer Webseite, inhaltlich verwandte Webseiten, verlinkte Webseiten etc.) des World Wide Webs sowie einzelner Webseiten verfügbar. Der *Alexa Thumbnail Service* (Preis: 0,20 $ je 1.000 Thumbnails) ermöglicht den Abruf und die Einbindung einer Thumbnail-Vorschau nahezu beliebiger Webseiten in eigene Anwendungen oder Webseiten.

8 Für eine detaillierte Beschreibung der Funktionalitäten, Schnittstellen sowie Kosten der einzelnen Services des Amazon-Webservice-Stack vgl. im Folgenden: http://www.Amazon.com/aws und http://developer.Amazonwebservices.com.

Die Schicht **People as a Service** weist nur einen einzigen Service auf, den *Mechanical Turk Service*. Der Mechanical Turk Service (Preis: 0,051 $ bis ca. 0,50 $ pro Aufgabe) ist ein Marktplatz für Aufgaben, die menschliche Fähigkeiten oder Intelligenz benötigen. Es gibt eine Vielzahl von Aufgaben, die ein Mensch wesentlich effizienter erfüllen kann als ein Computer, wie z. B. Bilderkennung, Übersetzungen oder Sprache in Text transformieren. Mittels des Mechanical Turk Service können Unternehmen oder Personen solche Aufgaben (sog. Human Intelligence Tasks) über Webserviceschnittstellen direkt aus ihren Anwendungen heraus in den Marktplatz einstellen und verwalten sowie die späteren, durch Menschen erbrachten Ergebnisse wieder in ihre Anwendung integrieren. Amazon spricht in diesem Zusammenhang von einer „API to human processing power“[9] oder auch von „artificial artificial intelligence“[10].

3 Betrachtung ausgewählter Services

Im Folgenden werden ausgewählte Services des Amazon-Webservice-Stack exemplarisch aus technischer, ökonomischer und sozialer Sicht analysiert. Als Beispiele werden der E-Commerce-Service, die Services Simple Storage und Elastic Compute Cloud sowie der Mechanical Turk Service herangezogen.

3.1 E-Commerce-Service (ECS)

Der Amazon-E-Commerce-Service war der erste von Amazon zur Verfügung gestellte Webservice. Im Vergleich zu anderen Services des Amazon-Webservice-Stack (z. B. Mechanical Turk, S3 oder EC2) handelt es sich dabei um einen klassischen Webservice im Sinne der Definition des W3C (vgl. Kapitel 1). Er ermöglicht, die Funktionalitäten und Daten der Amazon-Plattform über eine standardisierte XML-Schnittstelle zu nutzen und in beliebige eigene Webseiten oder Anwendungen zu integrieren. Der Service richtet sich in erster Linie an drei Zielgruppen:

- Amazon-Partner (sog. Associates) können mittels des Service Amazon-Inhalte effektiver in ihre eigenen Webseiten integrieren.
- Händler, die die Amazon-Plattform für den Verkauf eigener Artikel nutzen, können über den Service Produktdaten sowie Bestände verwalten.
- Entwickler können durch den Service Amazons Daten und/oder Funktionalitäten in eigene Anwendungen integrieren.

Mittels des E-Commerce-Service lässt sich die komplette Funktionalität der Amazon-Plattform nachbilden. Im Einzelnen handelt es sich dabei um: ausgiebige Such- und Navigationsmöglichkeiten innerhalb des Amazon-Produktkatalogs, detaillierte Informationen zu allen bei Amazon gelisteten Artikeln (inkl. Artikeln von Drittanbietern), aktuelle Preis- und Verfügbarkeitsinformationen, Artikelabbildungen in verschiedenen Größen, Informationen über Verkäufer, alle zu einem Artikel existierenden Kundenrezensionen, Verweise zu ähnlichen Artikeln („Kunden die

---

9 Barr (2007).
10 Arrington (2005).

diesen Artikel gekauft haben, haben auch folgende Artikel gekauft“) sowie Zugriff auf den Amazon-Einkaufswagen und die -Wunschliste.

Der E-Commerce-Service ist der einzige kostenlose Service im Amazon-Webservice-Stack. Die Motivation dahinter erklärt sich folgendermaßen: Je mehr Partner den Service in ihre Webseiten oder Anwendungen integrieren und somit bisher unerreichte Kunden anziehen, desto mehr Umsatz wird letztendlich für Amazon generiert. Denn die eigentliche Leistungserbringung, d. h. die Lieferung des gewünschten Artikels, sowie die Abwicklung der Bezahlung werden stets von Amazon erbracht. Die Partner kommen nur in der Vorkaufsphase, insb. in der Anwerbung des Kunden, ins Spiel. Alle nachgelagerten Aktivitäten verbleiben bei Amazon. Amazon gewährt den Partnern im Gegenzug eine Provision von bis zu 10 % auf durch den Service generierte Umsätze. Mit diesem Geschäftsmodel geht Amazon natürlich auch ein gewisses Risiko ein. Es besteht die Gefahr, dass irgendwann die Webseiten und Anwendungen der Partner beliebter werden als die eigentliche Amazon-Plattform. Dann tritt ein so genannter Kannibalisierungseffekt ein: Kunden, die normalerweise direkt auf der Amazon-Plattform einkaufen würden, nutzen die Webseiten und Anwendungen der Amazon-Partner, wodurch Amazon unnötige Provisionen an diese auszahlen muss.



Abbildung 2: TVmojo.com

Es existieren zahlreiche innovative Anwendungen, die auf dem Amazon-E-Commerce-Service basieren. *TVmojo.com* (vgl. Abbildung 2) ist ein Onlineshop für TV- und HIFI-Artikel.[11] Es handelt sich dabei um ein komplett virtuelles Unternehmen, das weder eigene Bestände führt noch eine logistische Infrastruktur (wie z. B. Lagerhäuser) besitzt. Die komplette Leistungserbringung

---

11 Vgl. im Folgenden Barr (2007).

wird durch Amazon übernommen. Selbst die TVmojo.com-Webseite läuft auf der IT-Infrastruktur von Amazon (EC2 und S3). Aus Anbietersicht liegt die Motivation von TVmojo.com in der Möglichkeit, einen auf ein Thema spezialisierten, aber zugleich sehr umfangreichen Onlineshop ohne großes Startkapital und ohne jegliche Fixkosten aufbauen zu können. Aus Nutzersicht liegt der Nutzen von TVmojo.com in der bewussten Begrenzung und bedarfsgerechten Aufbereitung des teilweise überwältigenden Artikelsortiments von Amazon. Zugleich minimiert die Tatsache, dass Amazon für die komplette Leistungserbringung sowie Zahlungsabwicklung verantwortlich ist, das Risiko für Anbieter sowie Nutzer.

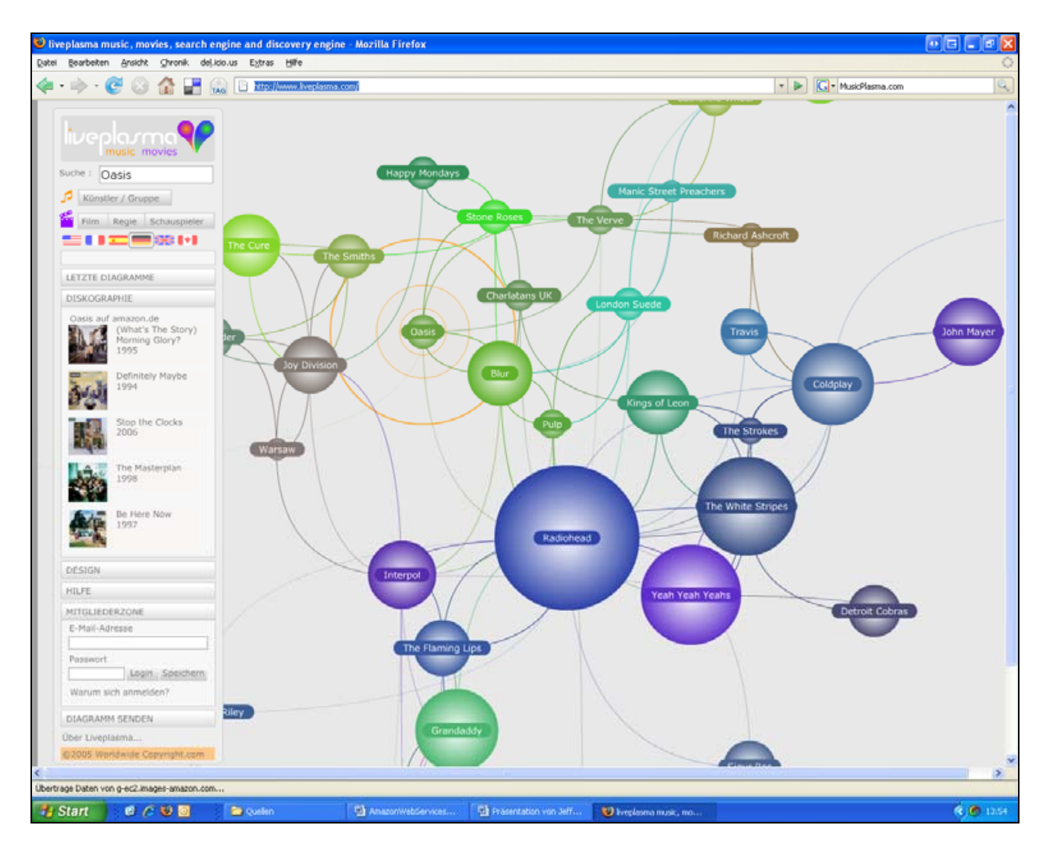


Abbildung 3: Liveplasma.com

Ein weiteres interessantes Beispiel, das die Mächtigkeit des E-Commerce-Service verdeutlicht, ist *liveplasma.com*.[12] Die Webseite bietet eine alternative Benutzeroberfläche für das Navigieren und Suchen innerhalb der Kategorien Musik und Film des Amazon-Produktkatalogs. Der Produktkatalog wird grafisch als baumartiges Diagramm, einem sog. Hyperbolic Tree, visualisiert (vgl. Abbildung 3). Knoten des Baums symbolisieren einzelne Künstler, die Größe des Knotens korrespondiert mit seiner Popularität und die Farbe steht für einen bestimmten Stil. Zudem stellen die Kanten des Baums Verbindungen zwischen Künstlern her, die eine musikalische Ähnlichkeit repräsentieren sollen. Die Diagramme basieren im Wesentlichen auf drei vom Amazon-E-Commerce-Service gelieferten Informationen:

- *SalesRank*: Der Rang eines Artikels in Amazons Verkaufshitliste. Diese Information wird für die Bestimmung der Popularität eines Künstlers, also der Größe eines Knotens, genutzt.

12 Vgl. im Folgenden Barr (2007).

- *BrowseNode*: Die Kategorie eines Artikels innerhalb des Amazon-Produktkatalogs. Diese Information bestimmt den Stil eines Künstlers, also die Farbe eines Knotens.
- *SimilarProduct*: Zu einem vorgegebenen Artikel ähnliche Artikel. Diese Information basiert auf Amazons Collaborative Filtering System und wird für die Konstruktion der Kanten zwischen Künstlern genutzt.

Im Gegensatz zu Geschäftsideen wie TVmojo.com steht hinter liveplasma.com weniger eine ökonomische Motivation. Es handelt sich vielmehr um ein Experiment zur Navigation in großen Datenbeständen, bei dem insbesondere der Erlebnisfaktor im Vordergrund steht:[13] *„Liveplasma is a new way to broaden your cultural horizons according to your taste in music and movies"*.

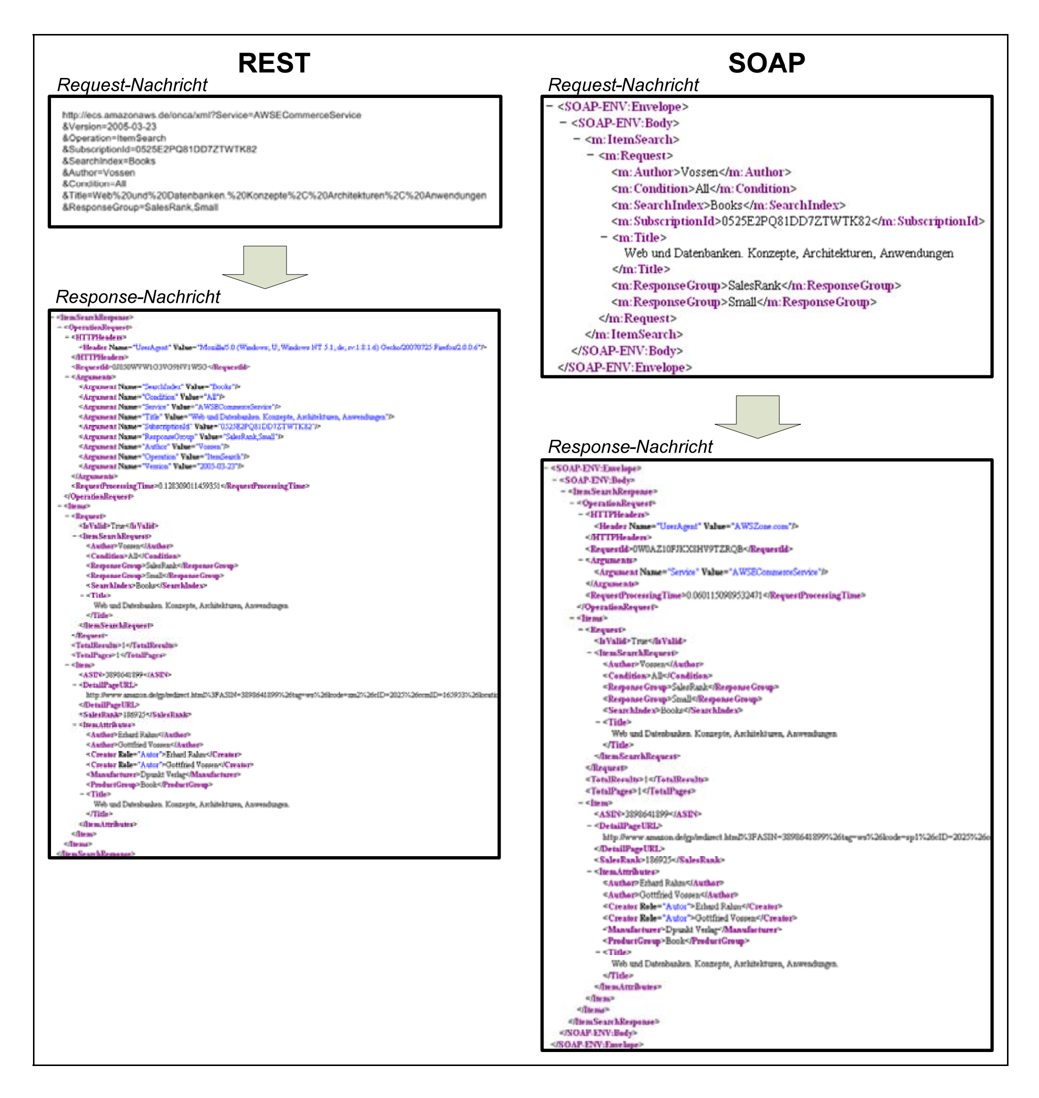


Abbildung 4: Vergleich von REST und SOAP

Aus technischer Sicht basiert die Nutzung des E-Commerce-Service, wie auch die Nutzung aller anderen Services des Amazon-Webservice-Stack, auf offenen und anerkannten Webservices (insb. WSDL, SOAP, REST, XML) und Internetstandards (insb. HTTP). Es stehen zwei Schnittstellen zur Verfügung: REST und SOAP.[14] Interessanterweise nutzen 85 % der Entwickler die

---

[13] Vgl. http://www.liveplasma.com.

[14] Für einen detaillierten Vergleich von REST und SOAP vgl. zur Muehlen, Nickerson, Swenson (2005).

REST-Schnittstelle, obwohl SOAP, nicht zuletzt durch die Unterstützung des W3C, in Wissenschaft und Praxis einen wesentlich höheren Bekanntheitsgrad erlangt hat.[15] Abbildung 4 zeigt die Unterschiede zwischen der Nutzung des E-Commerce-Service über REST und SOAP. Bei der Verwendung von REST wird die Request-Nachricht als einfache HTTP-Nachricht mit Parametern in der URL abgebildet. Im Fall von SOAP ist die Request-Nachricht ein separates XML-Dokument, das als Anlage an eine HTTP-Nachricht angehängt wird. Die Response-Nachrichten sind in beiden Fällen nahezu identische XML-Dokumente. Ein Erklärungsansatz für die Popularität von REST ist, dass kein extra SOAP-Werkzeug zur Generierung der (mitunter komplexen XML-basierten) Request-Nachrichten notwendig ist, sondern einfache URLs mit Parametern über beliebige HTTP-fähige Anwendungen erstellt und versendet werden können. Zudem ist REST aufgrund der engen Verwandtschaft zur Funktionsweise des World Wide Webs (WWW) intuitiver zu nutzen.

3.2 Simple Storage Service (S3) und Elastic Compute Cloud (EC2)

Klassische Webservices, wie beispielsweise der Amazon-Commerce-Service, kapseln die Funktionalität einer Softwarekomponente und machen sie über eine wohl definierte Schnittstelle verfügbar. Hinter den Amazon-Services S3 und EC2 steht eine andere Idee. Sie stellen Speicherplatz bzw. Rechenkapazität der Amazon Cloud über eine Webservice-API zur Verfügung.

Der S3-Service ermöglicht das Schreiben und Lesen von Datenobjekten nahezu unbegrenzter Größe (bis zu 5 GB) zu jeder Zeit und von jedem Ort des Internets aus. Technisch gesehen entspricht S3 einer über das Internet verteilten und zugreifbaren Hash Table, in der Objekte als Tupel von ID und Wert abgelegt werden.[16] Zudem besitzen sie neben den eigentlichen Daten Metadaten, wie z. B. Datum der letzten Änderung, Dateiformat oder selbst definierte Stichworte. Objekte sind in so genannten Buckets organisiert. Diese simplen Container dienen der Strukturierung und Adressierung von Objekten und stellen zudem die Basis für die Verwaltung von Zugriffsrechten sowie die Abrechnung von Kosten für Speicherplatz und Datentransfer dar. Das Senden von Objekten an S3 erfolgt, wie bei allen anderen Amazon-Webservices, wahlweise über eine SOAP- oder REST-Schnittstelle. Das Lesen von Objekten ist zudem über das offene, Peer-to-Peer-Protokoll BitTorrent möglich.

Ein kritischer Erfolgsfaktor eines jeden Systems zur Speicherung von Daten, der im Folgenden genauer untersucht werden soll, ist Datensicherheit.[17] Datensicherheit lässt sich in die Aspekte Vertraulichkeit, Integrität sowie Verfügbarkeit untergliedern. *Vertraulichkeit* wird beim S3-Service durch die Nutzung von Verschlüsselung und Zugriffskontrolle ermöglicht. Die Verschlüsselung der eigentlichen Datenobjekte wird durch S3 nicht unterstützt, kann jedoch leicht erreicht werden, indem die Objekte vor dem Senden an S3 verschlüsselt bzw. nach dem Empfangen von S3 wieder entschlüsselt werden. Objekt-IDs sowie Metadaten können verschlüsselt werden, indem die Kommunikation mit S3 über SSL und HTTPS abgewickelt wird. Die Zugriffskontrolle wird über eine Access Control List (ACL) auf der Ebene einzelner Buckets und User (d. h. S3-User) verwaltet. Zur Authentifizierung nutzt Amazon ein einfaches Verfahren auf Basis zweier Schlüssel: Die Access Key ID wird zur eindeutigen Identifikation eines S3-Users ver-

---

15 Vgl. O'Reilly (2003).

16 Vgl. Garfinkel (2007), S. 7.

17 Vgl. im Folgenden Garfinkel (2007), S. 7 ff.

wendet. Der Secret Access Key wird mit einem speziellen Algorithmus (SHA1-HMAC) für die Signierung von Request-Nachrichten verwendet. Der Server, dem die Secret Access Keys aller Nutzer bekannt sind, signiert eingehende Nachrichten mit eben dem gleichen Algorithmus und vergleicht die entstehende Werte, um die Authentizität des Nutzers zu überprüfen. Während dieses Verfahren im Allgemeinen als schnell, effizient und sicher angesehen wird, besteht die große Schwachstelle in der Tatsache, dass beide Schlüssel auf der AWS-Webseite abrufbar sind. Folglich kann jeder, der den Amazon-Login eines S3-Users kennt, in den Besitz der Schlüssel kommen und hat dann vollen Zugriff auf alle auf S3 abgelegten Datenobjekte. Zudem kann das Passwort der AWS-Webseite über eine übliche „Passwort vergessen?"-Funktion zurückgesetzt werden, wodurch jeder mit Zugriff auf das E-Mail-Postfach eines S3-Users auch an dessen Daten auf S3 gelangen kann. Zur Sicherstellung der *Integrität* während des Datentransports errechnet S3 beim Schreiben eines Objektes eine Prüfsumme (mittels des MD5-Algorithmus) und sendet diese nach der Übertragung als Antwort an den Client zurück. Dieser kann die Prüfsumme mit einer selbst errechneten Prüfsumme des korrespondierenden lokalen Objektes vergleichen, um festzustellen, ob die Übertragung fehlerfrei war bzw. nicht manipuliert wurde. Die Sicherung der Integrität während der langfristigen Speicherung von Daten wird nicht unterstützt. Der dritte Aspekt betrifft die *Verfügbarkeit*. Kurzfristige Verfügbarkeit wird durch die Erreichbarkeit des S3-Service repräsentiert. Tests haben gezeigt, dass übliche Verlustraten von Nachrichten zwischen 0,1 % und 1 % liegen.[18] Nach einer Wiederholung liegt die Verfügbarkeit bei durchschnittlich 99,55 % und spätestens nach der zweiten Wiederholung einer Nachricht konnten keine Datenverluste mehr festgestellt werden.[19] Diese Ergebnisse scheinen aufgrund der Tatsache, dass Amazon S3 auch für den Betrieb der eigenen Anwendungen nutzt, nicht verwunderlich und korrespondieren mit der von Amazon angegebenen Verfügbarkeitsrate von 99,99 %. Über die langfristige Verfügbarkeit von Daten auf S3 macht Amazon keine konkreten Angaben, behauptet lediglich, dass die Daten auf mehreren Speichermedien in verteilten Datencentern gespeichert werden. Jedoch existiert kein Backup-Service, über den Daten im Falle einer versehentlichen Löschung wiederhergestellt werden können. Neben den erwähnten Schwachstellen, insbesondere bei der Authentifizierung, stellt das Fehlen eines vertraglich geregelten Service Level Agreements (SLA) ein weiteres ernstzunehmendes Risiko für den professionellen Einsatz des S3-Service dar.

Der EC2-Service ermöglicht es, die Rechenkapazität der Amazon Cloud in Form virtueller Maschinen bedarfsweise zu beziehen. Jede dieser virtuellen Maschinen entspricht einem 1,7-GHz Intel Xenon-Server mit 1,25 GB Arbeitsspeicher. Persistenter Speicherplatz kann über S3 bezogen werden. Nutzer können auf die virtuellen Maschinen zugreifen, indem sie (über von Amazon zur Verfügung gestellte Werkzeuge) ein so genanntes Amazon Machine Image (AMI) erstellen und dieses auf S3 ablegen. Ein solches Image enthält neben dem eigentlichen Server (i. d. R. RedHat Linux Fedora Core) eine Firewall, die auszuführenden Anwendungen, alle nötigen Konfigurationen sowie die Daten, auf denen die Anwendungen arbeiten sollen. Über eine Webservice-API können die Images im Anschluss von externen Anwendungen aus gestartet, kontrolliert und beendet werden. EC2 ist skalierbar. Nutzer können beispielsweise eine Maschine für einen ganzen Monat oder 30 Maschinen für einen Tag oder 720 Maschinen für eine Stunde nutzen und zahlen dafür jeweils 72 $. Das Hinzu- oder Abschalten einzelner Maschinen

---

[18] Vgl. im Folgenden Palankar et al. (2007), S. 4.

[19] Stark schwankende Ergebnisse haben die Tests jedoch in punkto Datentransferrate ergeben, die sich je nach Kontinent zwischen 100 KB/s und 2 MBit/s bewegt.

über die EC2 API nimmt nur wenige Minuten in Anspruch, sodass die Rechenkapazität on-demand variiert werden kann. EC2 entspricht somit einer kommerziellen Form des Grid Computings, die nicht mehr nur für große Forschungseinrichtungen zugänglich ist, sondern von jedem, der im Besitz einer Kreditkarte ist, zu relativ niedrigen Kosten in Anspruch genommen werden kann. Typische Anwendungen kommen aus den Bereichen Datenanalyse, Simulation, 3D-Berechnung oder Testverfahren. Insbesondere das gerade eingeführte Konzept der Paid AMI macht den Service auch für einzelne Entwickler oder kleine Unternehmen interessant. Paid AMIs ermöglichen es, ein Image mit interessanten Anwendungen zusammenzustellen (in Amazons Worten: „a business in a box") und dieses anderen EC2-Usern im Sinne des Software as a Service (SaaS)-Konzeptes zu vermieten.

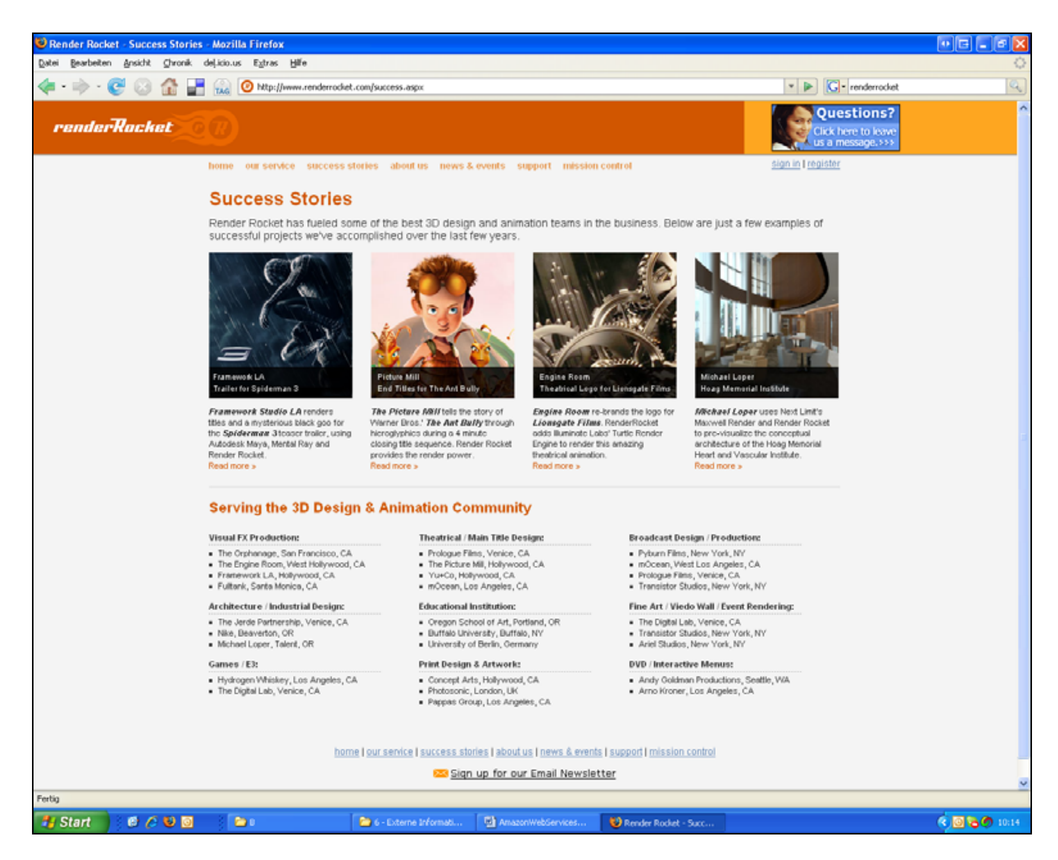


Abbildung 5: RenderRocket.com

Durch die Kombination von S3 und EC2 ist eine Reihe hoch innovativer Anwendungen entstanden: Ein Beispiel ist *RenderRocket* (vgl. Abbildung 5). RenderRocket erstellt professionelle 3D-Animationen für die Filmindustrie, Computerspieleentwickler, Designer, Architekten und viele mehr. Die Kunden von RenderRocket brauchen keine hochleistungsfähige Hard- oder Software, sondern können alle Dienste nach Bedarf über das Web in Anspruch nehmen. Noch interessanter ist jedoch die Tatsache, dass auch RenderRocket keine eigene IT-Infrastruktur besitzt, sondern die für das 3D-Renedering benötigten enormen Rechen- und Speicherkapazitäten je nach Auftragslage über die Amazon-Webservices bezieht. Das Beispiel verdeutlicht, wie durch die Amazon-Webservices das Start-up-Unternehmen RenderRocket eine innovative Geschäftsidee ohne große Vorabinvestitionen verwirklichen konnte, die es kleinen und kleinsten Teams erlaubt, ihre Ideen in einer vorher nicht möglichen Professionalität umzusetzen. Mittlerweile werden sogar Teile bekannter Hollywood-Blockbuster wie Spiderman 3 über RenderRocket animiert.

3.3 Mechanical Turk Service

Der Amazon Mechanical Turk Service ist weder mit dem klassischen Bild von Webservices (wie bspw. dem E-Commerce-Service) noch mit Infrastrukturservices wie S3 oder EC2 vergleichbar. Hier wird das Konzept des Service quasi umgedreht. Es wird nicht die Funktionalität von Hard- oder Software gekapselt und anderen Anwendungen oder Nutzern zur Verfügung gestellt, sondern menschliche Arbeitskraft und Intelligenz wird gekapselt und verfügbar gemacht. Abbildung 6 verdeutlicht die Funktionsweise des Service. Ein so genannter Requester erstellt eine Aufgabe (einen sog. Human Intelligence Task, HIT), die für einen Computer nur schwer zu erledigen ist, aber für einen Menschen kein großes Problem darstellt. Typische Beispiele sind Bild-, Text- oder Spracherkennung sowie die Bewertung von Inhalten oder auch Umfragen. Die Aufgabe kann neben einer Arbeitsanweisung und Vergütung auch benötigte Qualifikationen und Anlagen (z. B. Bilder, Audio, Textdokumente) enthalten. Der HIT wird auf der Mechanical Turk Webseite eingestellt, wo er durch menschliche, so genannte Worker gefunden werden kann. Hat ein Worker einen HIT angenommen und bearbeitet, so stellt er die Ergebnisse wiederum auf der Mechanical Turk-Webseite ein, von wo aus sie vom Requester abgerufen werden können. Nachdem der Requester sich von der Qualität der Arbeit überzeugt hat, veranlasst er die Überweisung der vorher festgelegten Vergütung, die in der Regel zwischen 0,05 $ und 0,50 $ liegt.

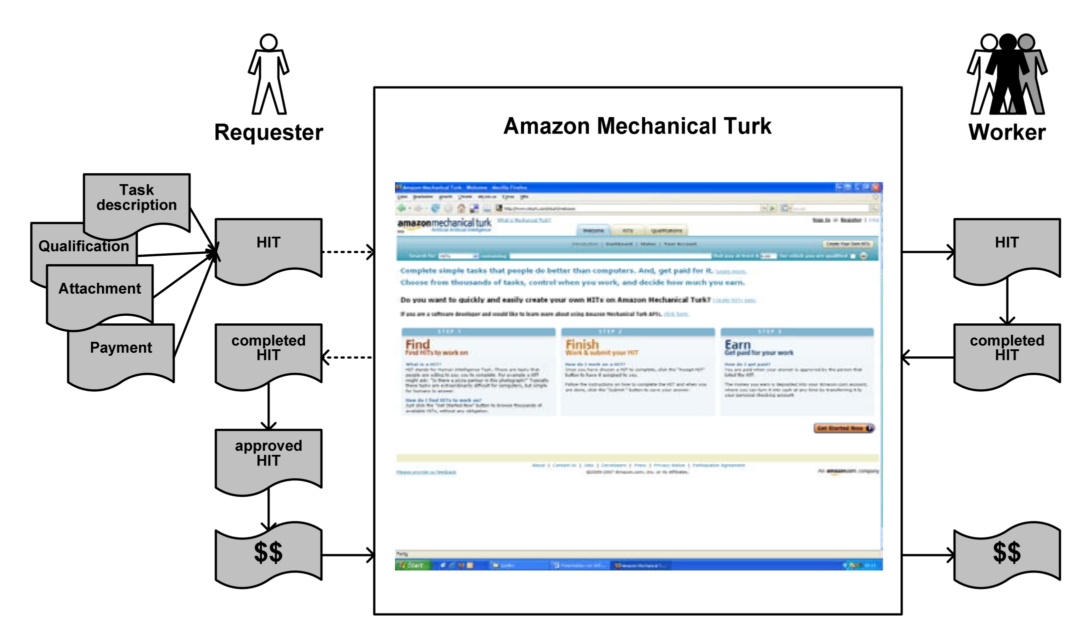


Abbildung 6: Funktionsweise des Mechanical Turk Service

Dadurch, dass das Einstellen der HITs und Abrufen der Ergebnisse über Webserviceschnittstellen erfolgen kann, ermöglicht der Mechanical Turk Service, eine skalierbare Menge an anonymer menschlicher Arbeitskraft und Intelligenz automatisiert in eine Softwareanwendung zu integrieren. Die Entkopplung von Requester und Worker ist dabei ein gewollter Effekt. Für Requester reduzieren sich dadurch die die eigentliche Aufgabe umgebenden Transaktionskosten (z. B. für die Suche von Arbeitskräften, Vertragsgestaltung oder langfristige Vertragsbeziehungen) erheblich. Für Worker, die bspw. aufgrund einer Behinderung nur noch schwer einer normalen Arbeit nachgehen können, eröffnet der Service ähnliche Potenziale. Ein Beispiel ist Mitch Fernandez, ein 38jähriger US-Army-Invalide, der von zu Hause teilweise mehr als 100 $ pro

Woche über den Mechanical Turk Service verdient.[20] Fernandez schreibt: *„I do this primarily for the money, but I also view it as a form of therapy to get me used to working again“.*[21] Jedoch findet die Idee, dass ein Computer Menschen vorgibt, welche Aufgaben sie zu erledigen haben, natürlich nicht nur Anhänger: *„[…] scary because I can't get the Matrix we-are-all-plugged-into-a-machine vision out of my head“*[22] oder *„There is something a little disturbing about a billionaire like Bezos dreaming up new ways to get ordinary folk to do work for him for pennies“*[23].

Ein amüsantes Beispiel für die Nutzung des Mechanical Turk-Service in Kombination mit anderen Amazon-Webservices ist die Anwendung *WhatsThatTune?*. Die Anwendung ermöglicht es, ein Lied über ein Mikrofon einzusingen, welches dann als MP3 auf dem S3-Service abgelegt wird. Dann wird ein HIT mit einem Link auf die MP3 und der Aufforderung, den Namen des Liedes zu nennen, auf der Mechanical Turk-Webseite eingestellt. Worker, die das Lied erkennen, können den Künstler und Liedtitel eingeben. Das Programm ermittelt im Anschluss mittels des E-Commerce-Service entsprechende CDs aus Amazons Produktkatalog und präsentiert diese dem Nutzer.

4 Ausblick

Die Erläuterungen ausgewählter Services sowie die realen Anwendungsbeispiele haben die technischen, ökonomischen sowie sozialen Potenziale des Amazon-Webservice-Stack verdeutlicht. Aus technischer Sicht hat Amazon bewiesen, dass auch umfangreiche und zuverlässige Anwendungen auf Basis der Webservicetechnologie realisiert werden können. Aus ökonomischer Sicht ist die Nutzung der Amazon-Webservices in vielen Bereichen hoch interessant. Ob sich das Geschäftsmodell jedoch auch für Amazon lohnt oder nur ein Experiment ist, bleibt abzuwarten. Aus sozialer Sicht ist es Amazon hoch anzurechnen, dass es wagt, seine Infrastruktur sowie Funktionalitäten und Daten, die einen komparativen Konkurrenzvorteil gegenüber Wettbewerbern darstellen und in die Unmengen von Geld und Know-how geflossen sind, quasi für jedermann zu öffnen. Gespannt ist abzuwarten, wie andere Softwaregiganten, allen voran Google und Yahoo (denen man solche Initiativen bestimmt eher zugetraut hätte), reagieren.

Literaturverzeichnis

Arrington, M.: Amazon finally reveals itself as the Matrix. 2005. http://www.techcrunch.com/2005/11/04/Amazon-finally-shows-itself-as-the-matrix/, Abrufdatum: 2007-09-05.

Arrington, M.: Interview with Jeff Bezos. 2006. http://www.talkcrunch.com/2006/11/14/interview-with-jeff-bezos/, Abrufdatum: 2007-09-05.

Barr, J.: 5 Minute AWS Presentation. 2007. http://aws.typepad.com/aws/2007/01/5_minute_aws_pr.html, Abrufdatum: 2007-09-05.

---

20 Vgl. Pontin (2007).
21 Pontin (2007).
22 Arrington (2005).
23 Pontin (2007).

Garfinkel, S.: Commodity grid computing with Amazon's S3 and EC2, in: Login, 32 (2007) 1, S. 7-13.

Haas, H.; Brown, A.: Webservices Glossary. W3C Working Group Note. 2004. http://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/, Abrufdatum: 2007-09-05.

Hof, R. D.: Jeff Bezos' Risky Bet, in: Business Week (online). 2006-11-13. http://www.businessweek.com/magazine/content/06_46/b4009001.htm, Abrufdatum: 2007-09-05.

Iskold, A.: Amazon Rolls Out its Visionary WebOS Strategy. 2006. http://www.readwriteweb.com/archives/Amazon_webos.php, Abrufdatum: 2007-09-05.

O'Reilly, T.: REST vs. SOAP at Amazon. 2003. http://www.oreillynet.com/pub/wlg/3005. Abrufdatum: 2007-03-09.

Palankar, M.; Onibokun, A.; Iamnitchi, A.; Ripeanu, M.: Amazon S3 for Science Grids: a Viable Solution? In: Poster, 4th USENIX Symposium on Networked Systems Design & Implementation, Cambridge, 2007.

Pontin, J.: Artificial Intelligence, with Help from Humans, in: The New York Times (online). 2007-03-25. http://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html?ex=1332475200&en=cd1ce5d0bee647d5&ei=5088. Abrufdatum: 2007-09-05.

Tapscott, D.; Williams, A. D.: Opening Up to Collaboration, in: Business Week (online). 2007-03-09. http://www.businessweek.com/innovate/content/mar2007/id20070309_322223.htm, Abrufdatum: 2007-09-05.

zur Muehlen, M.; Nickerson, J. V.; Swenson, K. D.: Developing Webservices Choreography Standards – The Case of REST vs. SOAP, in: Decision Support Systems 40 (2005) 1, S. 9-29.

P2P und VoIP

Christian Hermanns

1 Einleitung

Der Begriff Peer-to-Peer (P2P) beschreibt ein Netzwerk, in dem gleichberechtigte Knoten (Peers) einander Ressourcen wie Informationen, Speicherplatz, Rechenzeit oder Bandbreite zugänglich machen. Im Unterschied zum zentralisierten Client-Server-Modell gibt es in einem reinen P2P-Netzwerk keine dedizierten Client- oder Serverknoten, da jeder Peer gleichzeitig als Client und als Server fungiert. Die Peers können daher direkt miteinander kommunizieren, ohne auf die Vermittlung eines zentralen Servers zurückzugreifen.

Die Musiktauschbörse Napster, die auf dem Höhepunkt ihrer Popularität, im Februar 2001, 26,4 Mio. Benutzer verzeichnete, demonstrierte erstmals neue Möglichkeiten und Potenziale der P2P-Technologie. Heutzutage gibt es viele Anwendungsbereiche, in denen P2P-Anwendungen erfolgreich eingesetzt werden. Beispielsweise meldet der Internettelefoniedienst Skype [18] gegenwärtig über 220 Mio. Benutzer, von denen durchschnittlich 9 Mio. online sind. Eines der populärsten File-Sharing-Netzwerke, BitTorrent [3], ist verschiedenen Untersuchungen zufolge für 18 bis 35 % des gesamten Datenverkehrs im Internet verantwortlich [8]. Die Berkeley Open Infrastructure for Network Computing (BOINC) [22] ist eine Grid Computing-Plattform, die Rechner über das Internet zu einem virtuellen Supercomputer zusammenschließt. Die Plattform wurde ursprünglich für das SETI@home-Projekt entwickelt, betreibt gegenwärtig jedoch auch weitere Anwendungen aus den Bereichen Biologie, Medizin, Physik, Astronomie, Mathematik und Klimaforschung. BOINC verzeichnet gegenwärtig ca. 240.000 aktive Nutzer, die zusammen eine durchschnittliche Rechenleistung von ca. 515 TFLOPS erbringen und damit die Leistung des schnellsten Supercomputers Blue Gene [6] (360 TFLOPS) deutlich übertreffen.

Auch im Bereich der Forschung ist P2P vor allem in neuerer Zeit ein viel beachtetes Thema, so befassen sich laut CiteSeer [13] die meist zitierten Artikel aus den Jahren 2000 bis 2002 im Bereich der Informatik mit dem Thema P2P. Dabei ist das P2P-Konzept durchaus nicht neu: Die Server des Usenet oder die SMTP-Mailserver bilden untereinander dezentralisierte Netze aus gleichberechtigten Knoten, die direkte Kommunikationsbeziehungen zueinander unterhalten. Für die Nutzung beider Systeme ist jedoch ein entsprechender Client notwendig, um mit den P2P-Servern zu kommunizieren. Bei neueren P2P-Systemen sind die Zugangspunkte in der Regel keine Clients, sondern echte Peers. Möglich wurde diese Entwicklung durch stark gestiegene Rechenleistung und Bandbreite an den Zugangspunkten zum Internet, den Arbeitsplatzrechnern.

Im Rahmen dieses Beitrags wird ein Überblick über die technischen, ökonomischen und sozialen Aspekte von P2P-Systemen gegeben. Dieser Einleitung folgt im nächsten Kapitel eine Vorstellung der Anwendungsbereiche und Eigenschaften von P2P-Anwendungen. In Kapitel 3 werden die technischen Grundlagen von P2P-Systemen näher erläutert. Einen Überblick über die VoIP-Technologie gibt Kapitel 4. Die Kapitel 5 und 6 beschäftigen sich mit den ökonomischen

und sozialen Aspekten des P2P. Die Arbeit endet mit einer Zusammenfassung und einem Ausblick auf zukünftige Entwicklungen der P2P-Technologie.

2 P2P-Anwendungen

Der folgende Abschnitt beschreibt Anwendungsmöglichkeiten und Einsatzbereiche für die P2P-Technologie (Abschnitt 2.1) und zeigt besondere Eigenschaften von P2P-Systemen auf (Abschnitt 2.2).

2.1 Anwendungsbereiche

P2P-Systeme lassen sich anhand ihres Anwendungsbereiches in die drei Klassen *Verteiltes Rechnen*, *File Sharing* sowie *Kommunikation und Kollaboration* unterteilen. Im Folgenden werden diese Klassen ausführlicher vorgestellt.

Verteiltes Rechnen

Beim verteilten Rechnen berechnen die miteinander kommunizierenden Prozesse einer verteilten Anwendung ein gemeinsames Ergebnis. In der Regel laufen die Prozesse dabei auf unterschiedlichen Rechnern, die über ein Netzwerk miteinander verbunden sind.

Mithilfe des verteilten Rechnens ist es möglich, besonders rechenintensive Aufgaben zu lösen. Ein Problem wird dafür von einem zentralen Server in kleine unabhängige Teilprobleme aufgespalten, die anschließend auf die angeschlossenen Rechner verteilt werden. Nachdem ein Teilproblem durch einen Rechner gelöst wurde, wird das Ergebnis an den zentralen Server zurückgeschickt. Damit ein Rechner Rechendienste übernehmen kann, benötigt er einen Client, der die Kommunikation mit dem Server übernimmt und nicht benötigte Rechenzeit für die Lösung der ihm zugewiesenen Teilprobleme verwendet.

Die Tatsache, dass ein solches System über einen zentralen Server verfügt und zwischen den Clients keine Kommunikation stattfindet, spricht dafür, dass es sich nicht um P2P-Anwendungen handelt. Da aber dennoch ein Großteil der Anwendungslogik, d. h. der eigentliche Problemlösungsalgorithmus, auf den Clients ausgeführt wird und diese zudem eine große Autonomie besitzen, wird diese Art von Anwendungen zu den P2P-Anwendungen gezählt [12, 1].

Projekte, die diese Systeme verwenden, sind z. B. SETI@home zur Suche nach Hinweisen auf außerirdisches intelligentes Leben in Radiosignalen oder Folding@home zur Simulation von Proteinfaltungen, mit denen der Verlauf einiger Krankheiten wie Krebs oder Alzheimer besser verstanden werden kann. Aufgrund der großen Anzahl tRechner steht diesen Projekten eine enorme Rechenleistung zur Verfügung. Wie in der Einleitung bereits erwähnt, übersteigt die Rechenleistung dieser virtuellen Supercomputer sogar die der besten Supercomputer. Dabei ist allerdings zu berücksichtigen, dass die Kommunikation der einzelnen Knoten untereinander beim realen Supercomputer um ein Vielfaches schneller ist. Verteiltes Rechnen eignet sich daher hauptsächlich für Probleme, deren Lösung einen hohen Rechenaufwand und einen niedrigen Kommunikationsaufwand erfordern.

File Sharing

File Sharing gehört zu den bekanntesten und verbreitetsten P2P-Anwendungen im Internet. Ein File-Sharing-System ist ein dezentraler Datenspeicher, in dem Dateien mit einem Algorithmus gesucht und heruntergeladen werden können. Ein Teilnehmer kann beliebige Dateien für andere zum Download bereitstellen und umgekehrt Dateien von anderen Teilnehmern herunterladen. Der große Vorteil des P2P-File-Sharings liegt in der guten Skalierbarkeit. Da häufig getauschte Dateien in der Regel redundant gespeichert sind und daher von mehreren Quellen gleichzeitig heruntergeladen werden können, kommt es auch bei einer großen Anzahl von Downloads derselben Datei selten zu Engpässen in der Bandbreite. Bekannte Beispiele für File-Sharing-Netzwerke sind Gnutella [4], BitTorrent [3] oder Freenet [21].

Kommunikation und Kollaboration

Kollaborative P2P-Anwendungen dienen dazu, die Zusammenarbeit zwischen Nutzern in Echtzeit zu erleichtern. Da P2P-Systeme auf die Erzeugung von dynamischen Netzwerken mit variierenden Teilnehmern ausgelegt sind, eignen sie sich in natürlicher Weise für kollaborative Anwendungen. Eine der ersten P2P-Anwendungen in diesem Bereich waren die Instant-Messaging(IM)-Systeme, mit denen Textnachrichten in Echtzeit zwischen Teilnehmern ausgetauscht werden konnten. Mittlerweile unterstützen neuere Anwendungen auch Sprach- und Videokommunikation sowie den direkten Austausch von Dateien. Das Skype-Netzwerk [18] ist eine solche Anwendung, wobei hier der Fokus auf Sprachdiensten liegt.

Die P2P-Kollaborationssoftware Groove [11] bietet einen gemeinsamen Arbeitsbereich, der neben den oben genannten Funktionen weitere Dienste wie gemeinsame Diskussionsforen, Kalender und Dateifreigaben unterstützt. Groove synchronisiert automatisch alle Änderungen an Elementen des Arbeitsbereiches zwischen den Teammitgliedern. Die Änderungen können auch offline vorgenommen werden – die Synchronisierung erfolgt dann, sobald sich das betroffene Teammitglied mit dem Netzwerk der übrigen Mitglieder verbindet.

Häufig verfügen kollaborative Anwendungen über eine Teilnehmerliste, die anzeigt, welche Mitglieder gerade online sind, d. h., zu welchen Mitgliedern eine Verbindung aufgebaut werden kann. Diese Funktion ist besonders für Dienste mit Echtzeitkommunikation wichtig. Zur Aktualisierung dieser Liste muss der Status sämtlicher Mitglieder ständig kontrolliert werden. Da dies in einem reinen P2P-Netzwerk einen großen Kummunikationsaufwand bedeuten würde, kommt für diese Aufgabe meist ein zentraler Server zum Einsatz, bei dem sich alle Mitglieder an- und abmelden.

2.2 Eigenschaften

P2P-Anwendungen bestehen aus einem Zusammenschluss aus gleichberechtigten Knoten zu einem Netzwerk und stellen damit einen Gegensatz zu den zentralisierten Client-Server-Architekturen dar. Aus dem Paradigma der Dezentralisierung ergeben sich für Anwendungen mit P2P-Architekturen einige grundlegende Eigenschaften, die sie von anderen Architekturen unterscheidet (vgl. [12]).

Dezentralisierung

In P2P-Systemen werden Informationen und Ressourcen nicht durch einen zentralen Server, sondern durch gleichberechtigte Peers bereitgestellt. Obwohl einige Probleme, wie z. B. die Suche nach Inhalten, Authentifizierung oder Autorisierung, in dezentralen Systemen schwieriger zu realisieren sind, zeichnen sie sich in der Regel durch eine effizientere Nutzung der Netzwerkressourcen aus (s. u.). In einem reinen P2P-System sind alle Peers vollkommen gleichartige Teilnehmer, die sowohl Client als auch Server sind.

Skalierbarkeit

Da in einem P2P-Netzwerk eine Gleichverteilung der Ressourcenauslastung angestrebt wird und die Ressourcen mit jedem weiteren Peer zunehmen, weisen P2P-Netzwerke in der Regel eine sehr gute Skalierbarkeit auf. Diese Eigenschaft unterscheidet sie deutlich von den zentralisierten Client-Server-Systemen, die eine schlechtere Skalierbarkeit aufweisen, da die Ressourcen, die ein Server einem einzelnen Client zur Verfügung stellen kann, mit zunehmender Anzahl der Clients sinken.

Aggregation von Ressourcen

Viele P2P-Anwendungen basieren auf dem Prinzip, dass sie die Ressourcen (Rechenleistung, Bandbreite, Speicherplatz) einzelner Peers aggregieren. Ein so entstehendes virtuelles System verfügt bei ausreichender Anzahl von Peers über eine große Menge der entsprechenden Ressource. Dieser Ansatz wird zum Beispiel von Systemen für verteiltes Rechnen oder File Sharing verfolgt.

Robustheit

Reine P2P-Netzwerke sind selbstorganisierend und besitzen keine zentrale Kontrollinstanz. Da die Organisation des Netzes durch die Gesamtheit der miteinander kommunizierenden Peers übernommen wird, gibt es keinen zentralen Knoten im Netzwerk, dessen Ausfall den Zusammenbruch des gesamten Netzwerkes zur Folge hätte.

Kosten

In einem zentralisierten Client-Server-System entfällt ein Großteil der Betriebskosten für die notwendigen Ressourcen wie Rechenleistung, Bandbreite oder Speicherplatz auf den zentralen Server. Bei P2P-Architektur werden diese Kosten auf alle Peers des Netzwerks verteilt, indem jeder einen Teil der benötigten Ressourcen beiträgt. Neben dieser *Kostenverteilung* ergibt sich aber häufig auch eine *Kostenreduktion*, wenn die Peers über freie Ressourcen wie ungenutzte Bandbreite oder Rechenleistung verfügen und diese dem System ohne zusätzliche Kosten zur Verfügung stellen.

Dynamik

Ein wichtiger Aspekt von P2P-Netzwerken ist deren große Dynamik. Ständig treten Peers dem Netz bei oder verlassen es. P2P-Anwendungen müssen entsprechende Mechanismen enthalten, um auf diese Dynamik reagieren zu können. Bei einer IM-Anwendung ist z. B. die Liste der erreichbaren Leute ständig zu aktualisieren oder beim verteilten Rechnen kann es vorkommen,

das Peers dass Netzwerk verlassen, bevor sie die Lösung zu einem zugeteilten Problem zurückgeschickt haben.

Anonymität

In einem Client-Server-System ist es schwierig, die Anonymität eines Clients zu gewährleisten, da ein Server in der Regel in der Lage ist, die mit ihm verbundenen Clients anhand ihrer IP-Adresse zu identifizieren. In einem dezentralisierten Netzwerk ist es für einen Peer sehr viel leichter, keine Informationen über die eigenen Aktivitäten preiszugeben. P2P-Netzwerke wie Freenet [21] besitzen eine Struktur, die speziell auf die Wahrung der Anonymität der Teilnehmer ausgelegt ist.

3 P2P-Netzwerkarchitekturen

Der folgende Abschnitt betrachtet die technischen Aspekte von P2P-Architekturen. Ausgehend von der Beschreibung eines reinen P2P-Netzwerks (Abschnitt 3.1) werden zwei Möglichkeiten aufgezeigt, mit denen die in der Regel sehr aufwendige Suche in P2P-Netzwerken optimiert werden kann (Abschnitt 3.2): Das Netzwerk kann entweder *zentralisiert* werden oder die gesuchten Inhalte können *strukturiert* gespeichert werden.

3.1 Aufbau eines reinen P2P-Netzwerks

Ein reines P2P-Netzwerk besteht ausschließlich aus gleichberechtigten Peers und steht daher im Gegensatz zum zentralisierten Client-Server-Modell. Die Abbildung 1a) zeigt den schematischen Aufbau eines solchen Netzes. Jeder Peer ist mit einer Teilmenge der übrigen Peers verbunden und übernimmt die gleichen Aufgaben. Auf ein reines P2P-Netzwerk treffen alle im Abschnitt 2.2 genannten Eigenschaften zu. In seiner ursprünglichen Form war z. B. das Gnutella-Netzwerk [4] ein solches vollkommen dezentralisiertes Netzwerk.

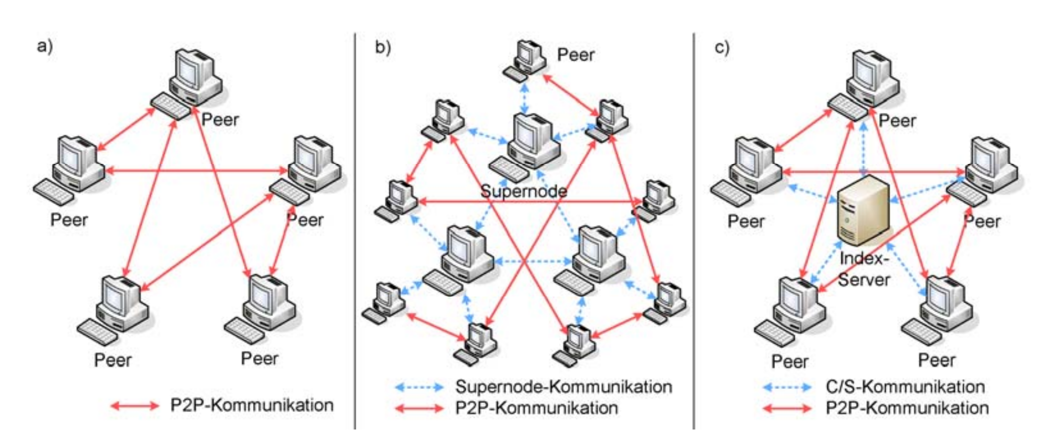


Abbildung 1: Aufbau eines reinen P2P-Netzwerks a), eines P2P-Netzwerkes mit Supernodes b) und eines P2P-Netzwerkes mit zentralem Index-Server c)

Der große Nachteil der reinen P2P-Netzwerke zeigt sich, sobald diese nach bestimmten Inhalten durchsucht werden müssen. Da sowohl das Netz selber als auch dessen Inhalte keinerlei Struktur aufweisen, müssen immer alle Knoten des Netzes durchsucht werden, um ein vollstän-

diges Suchergebnis zu erhalten. Um z. B. in einem File-Sharing-Netz eine Datei zu finden, sendet ein Peer eine Suchanfrage an alle mit ihm verbundenen Peers. Ein Peer, der eine Suchanfrage erhält, leitet diese wiederum an alle ihm bekannten Peers weiter, wobei Zyklen durch geeignete Techniken vermieden werden. Ist die Suche beim durchsuchten Peer erfolgreich, so wird dies dem suchenden Peer gemeldet. Die Suche endet, sobald alle Peers befragt wurden. Da durch diese Art der Suche die Ressourcen aller Peers beansprucht werden, nimmt die Ressourcenauslastung eines einzelnen Peers mit der Häufigkeit der Suchanfragen zu. Da davon auszugehen ist, dass die Anzahl der Suchanfragen mit der Anzahl der Peers steigt, skalieren reine P2P-Netze sehr schlecht, wenn es zu regelmäßigen Suchanfragen kommt.

Es gibt verschiedene Ansätze, diesem Effekt entgegenzuwirken, indem beispielsweise die maximale Suchtiefe einer Anfrage beschränkt wird. Dies hat jedoch zur Folge, dass das Suchergebnis nicht vollständig ist. Um eine effizientere Suche zu ermöglichen, ist es nötig, den Aufbau des Netzes zu ändern oder die gesuchten Inhalte zu sortieren. Beide Ansätze werden im Folgenden beschrieben.

3.2 Optimierung der Suche

Zentralisierung

Zentralisierung ist eine Möglichkeit, die Effizienz der Suche in einem P2P-Netzwerk zu verbessern. Dabei wird das reine P2P-Paradigma aufgebrochen, indem Knoten eingeführt werden, die einen größeren Aufgabenbereich haben als andere. Sie besitzen eine Liste über die Inhalte der mit ihnen verbundenen Knoten. Eine Suchanfrage muss dann nur noch an alle Superknoten gesendet werden. Folglich entspricht die durchschnittliche Anzahl der Konten, die einem Superknoten zugeordnet sind, in etwa dem Faktor, um den der Suchaufwand reduziert wird. Im Extremfall gibt es nur einen einzigen solchen Supernode. Da in diesem Fall eine Suchanfrage direkt durch den zentralen Index-Server beantwortet werden kann, ist der Suchaufwand auch bei steigender Knotenzahl konstant. Die Möglichkeiten, ein P2P-Netz zu zentralisieren, werden in den Abbildung 1b) und Abbildung 1c) dargestellt.

Ein Beispiel für eine P2P-Architektur mit zentralem Index-Server ist die Tauschbörse Napster. Alle Peers des Napster-Netzwerks registrieren sich mit einer Liste zur Verfügung gestellter Dateien beim Server. Eine Suchanfrage an den zentralen Server liefert eine Liste aller Peers, die die gesuchte Datei bereitstellen. Um die Datei herunterzuladen, werden die infrage kommenden Peers direkt kontaktiert. Eine Architektur mit Superknoten, die im Punkto Zentralisierung einen Kompromiss zwischen einem reinen P2P-Netzwerk und einem P2P-Netzwerk mit zentralem Index-Server darstellt, besitzt u. a. das Skype-Netzwerk.

Mit zunehmender Zentralisierung eines P2P-Netzwerkes steigt zwar die Geschwindigkeit von Suchanfragen, dafür werden jedoch einige positive Eigenschaften von reinen P2P-Netzen eingebüßt. Beispielsweise sinkt die Robustheit des Systems, da durch Ausfälle der zentralen Superknoten das Netzwerk teilweise oder auch vollständig zusammenbrechen kann. In besonderem Maße trifft dies zu, falls es nur einen zentralen Index-Server gibt. In diesem Fall ist auch zu berücksichtigen, dass die Skalierbarkeit des Netzwerkes eingeschränkt ist, da die Ressourcen des Servers an die Anzahl der Peers bzw. der Suchanfragen angepasst werden müssen.

Strukturierung

Neben der Hierarchisierung der Netzwerkstruktur ist die Strukturierung des Inhalts eine weitere Möglichkeit, die Sucheffizienz eines P2P-Netzwerks zu erhöhen. Der Vorteil dieses Ansatzes besteht darin, dass es keine Superknoten gibt, deren Ausfall die Robustheit des Netzwerkes einschränkt.

Die strukturierten P2P-Netzwerke zeichnen sich dadurch aus, dass die gesuchten Inhalte an bestimmten Knoten im Netz gespeichert werden, wodurch sich der Suchaufwand reduziert. Hierfür werden sowohl die Knoten als auch die Inhalte mit eindeutigen Schlüsseln, welche z. B. durch eine Hashfunktion berechnet werden, versehen. Die Idee ist, die Inhalte in Knoten zu speichern, deren Schlüssel dem Schlüssel der Inhalte möglichst ähnlich ist. Um einen solchen Knoten aufzufinden, besitzt jeder Knoten eine Routing-Tabelle, in der die Schlüssel und Adressen mehrerer Knoten gespeichert sind. Erhält ein Knoten eine Suchanfrage, zu einem Schlüssel den ähnlichsten Knoten zu finden, so leitet er diese Anfrage an den ähnlichsten Knoten in seiner Routing-Tabelle weiter und gibt das von diesem Knoten erhaltene Ergebnis zurück. Sollte der Knoten selbst der ähnlichste sein, so leitet er die Anfrage nicht weiter, sondern gibt die eigene Adresse als Ergebnis zurück.

Im Prinzip funktionieren die meisten strukturierten P2P-Netzwerke nach diesem Schema und variieren lediglich im Aufbau der Routing-Tabelle. Im Chord-Netzwerk [19] verwaltet z. B. jeder Knoten logN weitere Knoten in seiner Routing-Tabelle (wobei N der Gesamtzahl der Knoten entspricht), während die Anzahl der Knoten im CAN-Netzwerk [15] einem konstanten Wert entspricht. Weiterführende Untersuchungen der Suchproblematik in dezentralen P2P-Netzwerken finden sich u. a. bei [2, 1].

4 VoIP

Voice over Internet Protocol (VoIP) bezeichnet die Übertragung von Telefongesprächen über das Internet oder andere IP-basierte Netzwerke. Bei VoIP werden analoge Sprachsignale digitalisiert und in Form von Datenpaketen über ein IP-Netzwerk zum Empfänger übertragen. VoIP ist eine P2P-Technologie, da eine direkte Kommunikationsbeziehung zwischen beiden Gesprächspartnern aufgebaut wird.

Im Unterschied zu leitungsvermittelten Telefonnetzen wird bei VoIP für ein Telefongespräch keine dedizierte Leitung reserviert, weil die Daten paketbasiert durch ein IP-Netzwerk geleitet werden. Durch die Paketvermittlung der Daten bietet VoIP den Vorteil, dass vorhandene Netzwerkressourcen wesentlich effizienter genutzt werden können. Langfristig ist geplant, das Telefonnetz sogar vollständig durch die VoIP-Technologie zu ersetzen. Für Daten- und Sprachkommunikation gäbe es dann nur noch ein einziges Netz. VoIP bietet daher enorme Kostensenkungspotenziale.

Um das Telefonnetz ersetzen zu können, muss VoIP ähnliche Qualität und Zuverlässigkeit bieten wie das Telefonnetz. Die Anforderungen an die Güte eines Kommunikationsdienstes werden in der Regel durch die *Quality of Service (QoS)* beschrieben. In einem paketvermittelten Netzwerk wird die QoS im Wesentlichen durch die vier Faktoren *Verlässlichkeit*, *Verzögerung*,

Jitter und *Bandbreite* beschrieben [20]. Verlässlichkeit bezeichnet die Wahrscheinlichkeit, dass ein Paket den Empfänger erreicht, die Verzögerung die durchschnittliche Laufzeit eines Paketes, der Jitter die Standardabweichung der durchschnittlichen Laufzeit und die Bandbreite die maximal zustellbare Paketanzahl pro Zeiteinheit. Je höher die Verlässlichkeit, je kleiner die Verzögerung, je kleiner der Jitter und je größer die Bandbreite, desto größer die QoS.

In heutigen IP-Netzwerken werden die für VoIP notwendigen QoS zwar in der Regel erreicht, es ist jedoch schwer möglich, diese zu garantieren. Da auf das Routing der Pakete kein Einfluss genommen werden kann, ist es bei hohen Netzwerkauslastungen möglich, dass die QoS nicht eingehalten werden können. Um die QoS mit hinreichender Sicherheit und ökonomisch sinnvoll garantieren zu können, ist eine Priorisierung der Pakete notwendig. Dienste wie VoIP, die einen relativ hohen Anspruch an die QoS stellen, können dann priorisiert werden, um auch bei hohen Netzwerkauslastungen die geforderte QoS zu garantieren. Die Netzwerkarchitektur, die Sprach- und Datendienste über ein gemeinsames IP-Netzwerk abwickelt, wird auch als *Next Generation Network (NGN)* bezeichnet.

Für VoIP existieren zwei bedeutende internationale Standards, der H.323-Standard [7] der International Telecommunication Union (ITU) und das Session Initiation Protocol (SIP) [5] der Internet Engineering Task Force (IETF). Beide Standards definieren, wie eine Kommunikation zwischen zwei oder mehreren Teilnehmern *aufgebaut* und *durchgeführt* wird. Während beide Standards das *Real-time Transport Protocol (RTP)* [16] für die Durchführung der Kommunikation verwenden, ist der Aufbau einer Kommunikationsbeziehung unterschiedlich. SCHULZRINNE und ROSENBERG kommen zu dem Schluss, dass beide Standards in etwa den gleichen Funktionsumfang besitzen, wobei SIP weniger komplex ist, besser skaliert und gute Erweiterungsmöglichkeiten bietet [17]. Neben diesen beiden offenen Standards gibt es auch proprietäre Systeme wie Skype, die VoIP-Lösungen anbieten. Es ist fraglich, inwieweit sich diese in Zukunft gegenüber den offenen Standards behaupten können.

5 Ökonomische Aspekte von P2P

Die P2P-Technologie ermöglicht es durch ihren dezentralen Charakter, Kosten für benötigte Systemressourcen zu verteilen und zu reduzieren. Von dieser Kostenreduktion können sowohl die Anbieter als auch die Nutzer eines Dienstes profitieren. Beim verteilten Rechnen nutzen Projekte wie SETI@home ungenutzte Rechenleistung, die Peers kostenlos zur Verfügung stellen, und ersparen sich so teure Investitionen in zusätzliche Hardware. Ähnlich lassen sich durch File Sharing die Distributionskosten für digitale Inhalte reduzieren. Beispielsweise verwendet die Firma Blizzard das P2P-Netzwerk BitTorrent, um Updates des Spiels World of Warcraft kostengünstig an alle Abonnenten zu verteilen. Der Telefoniedienst Skype ist ein gutes Beispiel für eine Kostenreduktion beim Dienstnutzer, da Gespräche von Skype-Teilnehmer zu Skype-Teilnehmer kostenlos sind.

File-Sharing-Netzwerke werden häufig dazu gebraucht, auch urheberrechtlich geschützte Inhalte illegal auszutauschen. Inhaber des Urheberrechts bzw. des Nutzungsrechts können dadurch erheblich in ihren Interessen geschädigt werden. P2P-Netzwerke sind für das File Sharing besonders beliebt, da aufgrund ihres dezentralen Charakters eine Kontrolle der ausgetauschten Inhalte nur sehr schwer oder auch gar nicht möglich ist. Zudem mindert die Anonymität das Ri-

siko, erwischt zu werden und erhöht so die Bereitschaft zum illegalen Tausch geschützter Inhalte. Aus ökonomischer Sicht kann die P2P-Technologie daher auch Einnahmeverluste verursachen, indem sie die Zahlungsbereitschaft für den legalen Erwerb geschützter Inhalte reduziert.

6 Soziale Aspekte von P2P

Während P2P-Anwendungen für File Sharing oder verteiltes Rechnen in der Regel Netzwerke aus anonymen Peers bilden, gibt es auch P2P-Anwendungen, die die Bildung von privaten Communities unterstützen. Zu dieser Kategorie gehören die meisten P2P-Anwendungen aus dem Bereich der Kommunikation und Kollaboration. Diese Anwendungen besitzen üblicherweise eine Kontaktliste von Personen, mit denen über das Netzwerk kommuniziert werden kann. In diesem Fall bildet jeder Peer ein privates P2P-Netzwerk mit den Peers seiner Kontaktliste. Anwendungen für Instant Messaging und Sprach- und Videotelefonie sind Beispiele für solche privaten sozialen P2P-Netzwerke.

Darüber hinaus gibt es auch weiterführende Lösungen, die nicht nur die Kommunikation, sondern auch das File Sharing betreffen. Die Anwendungen Qnext [14] und Groove ermöglichen es, lokale Dateien für Kontakte freizugeben. Die freigegebenen Dateien erscheinen dann in einem virtuellen gemeinsamen Ordner, wo sie von autorisierten Kontakten heruntergeladen werden können, wenn der freigebende Peer online ist. Während Qnext eher für den Einsatz im privaten Bereich konzipiert ist, ist Groove für die Koordination von Arbeitsgruppen konzipiert. Groove bietet daher weiterführende Kollaborationsdienste, wie Kalender- oder Whiteboardfunktionen.

Der Nutzen von P2P-Netzwerken – insbesondere von den hier vorgestellten privaten P2P-Netzwerken – hängt oft in entscheidendem Maße von der Anzahl der Teilnehmer ab. Durch diesen Netzwerkeffekt [10] ist der Nutzen eines einzelnen Teilnehmers umso größer, je größer die Anzahl der Teilnehmer in einem P2P-Netzwerk ist. Darüber hinaus gibt es auch für P2P-Netzwerke eine kritische Masse, d. h. eine Mindestanzahl an Teilnehmern, die überschritten werden muss, damit der Nutzen den Aufwand für die Teilnahme an einem P2P-Netzwerk übersteigt.

7 Zusammenfassung und Ausblick

Im Rahmen dieser Ausarbeitung wurden die wesentlichen Anwendungsbereiche und Eigenschaften von P2P-Anwendungen vorgestellt. P2P-Anwendungen bilden ein dezentrales Netzwerk gleichartiger Knoten und unterscheiden sich damit grundlegend vom Client-Server-Paradigma. Zwar existieren internetbasierte P2P-Anwendungen bereits seit den Anfangszeiten des Internets, durch den Siegeszug des Internetbrowsers war jedoch lange Zeit die Client-Server-Architektur dominierend für verteilte Systeme. Erst die zunehmende Bandbreite und Rechenleistung hat den Betrieb der hier vorgestellten P2P-Systeme auf einem herkömmlichen Arbeitsplatzrechner ermöglicht und zu deren Erfolg beigetragen.

Bei der technischen Betrachtung von P2P-Systemen wurde die grundlegende Architektur und Funktionsweise eines reinen P2P-Systems beschrieben. Während vollkommen dezentrale Systeme sich durch effiziente Ressourcennutzung auszeichnen, sind Funktionen wie Authentifizie-

rung, Autorisierung oder Suche nur mit hohem Aufwand zu realisieren. Um dieses Defizit auszugleichen, kann das Netz entweder zentralisiert oder strukturiert werden.

Die VoIP-Technologie, die es ermöglicht, Sprach- und Videokommunikation paketvermittelt über ein IP-Netzwerk durchzuführen, basiert ebenfalls auf dem P2P-Prinzip und bietet im Vergleich zur herkömmlichen Telefonie enorme Kostenvorteile. Um für VoIP einen vergleichbaren QoS zu garantieren, müssen einzelne Dienste in IP-Netzwerken priorisiert werden können. Diese Anforderungen sollen durch das NGN erfüllt werden.

Aus ökonomischer Sicht bieten P2P-Netzwerke aufgrund ihrer effizienten Ressourcen Nutzung enorme Kostenvorteile. Einsparungen sind sowohl aufseiten des Anbieters als auch des Nutzers möglich.

P2P-Netzwerke können auch dazu dienen, Communities zu bilden. Dies sind abgeschlossene private Netzwerke die Freunden, Bekannten oder Arbeitskollegen eine gemeinsame Kommunikations- oder Arbeitsplattform bieten.

Aufgrund der gegenwärtigen Entwicklungen ist davon auszugehen, dass zukünftige Anwendungen im Bereich der P2P-Technologie sich insbesondere mit dem Streaming von multimedialen Inhalten befassen. Diese Anwendungen haben einen hohen Bedarf an Bandbreite und sind deshalb erst seit kurzem realisierbar. Neue Entwicklungen wie das P2P-TV Joost [9] werden durch die schnell zunehmenden Bandbreiten der Internetanschlüsse begünstigt. Des Weiteren bleibt es abzuwarten, ob sich eine Lösung für die zahlreichen Urheberrechtsverletzungen, welche durch P2P-Technologien begünstigt werden, findet.

Literaturverzeichnis

[1] S. Androutsellis-Theotokis, A survey of peer-to-peer file sharing technologies, www.cs.ucr.edu/~michalis/COURSES/179-03/p2psurvey.pdf, Abrufdatum: 29.08.2007.

[2] H. Balakrishnan, M. F. Kaashoek, D. Karger, R. Morris, I. Stoica, Looking up data in p2p systems, In Communications of the ACM, volume 46, pages 43-48, New York 2003.

[3] BitTorrent, Inc. Bittorrent. http://www.bittorrent.com, Abrufdatum: 29.08.2007.

[4] Gnutella, http://www.gnutella.com, Abrufdatum: 29.08.2007.

[5] M. Handley, H. Schulzrinne, E. Schooler, J. Rosenberg, Sip: Session initiation protocol, http://www.ietf.org/rfc/rfc2543.txt, Abrufdatum: 29.08.2007.

[6] IBM, Blue gene, http://www.research.ibm.com/bluegene, Abrufdatum: 29.08.2007.

[7] International Telecommunication Union, H. 323, http://www.itu.int/rec/T-REC-H.323/en, Abrufdatum: 29.08.2007.

[8] ipoque. P2p-studie 2006 – Kurzfassung. www.ipoque.com/pub/P2P-Studie-2006.pdf, Abrufdatum: 29.08.2007.

[9] N. V. Joost, Joost, http://www.joost.com/, Abrufdatum: 29.08.2007.

[10] M. L. Katz, C. Shapiro, System competition and network effects. In Journal of Economic Perspectives, pages 93-115, 1994.

[11] Microsoft, Groove, www.groove.net, Abrufdatum: 29.08.2007.

[12] D. S. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, Z. Xu, Peer-to-peer computing. www.hpl.hp.com/techreports/2002/HPL-2002-57R1.pdf, Abrufdatum: 29.08.2007.

[13] NEC, PennState, Citeseer, http://citeseer.ist.psu.edu/, Abrufdatum: 29.08.2007.

[14] Qnext Corp, Qnext, http://qnext.com/, Abrufdatum: 29.08.2007.

[15] S. Ratnasamy, P. Francis, M. Handley, R. Karp, S. Shenker, A scalable content-addressable network. In SIGCOMM ’01, pages 161-172, New York 2001.

[16] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, Rtp: A transport protocol for real-time applications, http://www.faqs.org/rfcs/rfc1889.html, Abrufdatum: 29.08.2007.

[17] H. Schulzrinne, J. Rosenberg. A comparison of sip and h.323 for internet telephony, www.cs.columbia.edu/~hgs/papers/Schu9807_Comparison.pdf, Abrufdatum: 29.08.2007.

[18] Skype Limited, Skype, http://www.skype.com, Abrufdatum: 29.08.2007.

[19] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, H. Balakrishnan, Chord: A scalable peer-to-peer lookup service for internet applications. In SIGCOMM ’01, pages 149-160, New York 2001.

[20] A. S. Tanenbaum, Computer Networks, 4. Aufl., Prentice Hall PTR, New Jersey 2003.

[21] The Free Network Projekt, Freenet. http://freenetproject.org/, Abrufdatum: 29.08.2007.

[22] Univerity of California, Boinc – berkeley open infrastructure for network computing, http://boinc.berkeley.edu/, Abrufdatum: 29.08.2007.

Arbeitsberichte des Kompetenzzentrums Internetökonomie und Hybridität

Grob, H. L. (Hrsg.), Internetökonomie und Hybridität – Konzeption eines Kompetenzzentrums im Forschungsverbund Internetökonomie, Nr. 1.

Brocke, J. vom, Hybride Systeme - Begriffsbestimmung und Forschungsperspektiven für die Wirtschaftsinformatik, Nr. 2.

Holznagel, D., Krone, D., Jungfleisch, C., Von den Landesmedienanstalten zur Ländermedienanstalt - Schlussfolgerungen aus einem internationalen Vergleich der Medienaufsicht, Nr. 3.

Zimmerlich, A., Aufderheide, D., Herausforderungen für das Wettbewerbsrecht durch die Internetökonomie, Nr. 4.

Ahlert, D., Evanschitzky, H., Erfolgsfaktoren des Multi-Channel-Managements, Nr. 5.

Holling, H., Freund, P. A., Kuhn, J.-T., Usability-Analysen von Wissensmanagementsystemen, Nr. 6.

Bröcher, J., Domain-Names und das Prioritätsprinzip im Kennzeichenrecht – Nochmals shell.de & Co., Nr. 7.

Trauten, A., Zur Effizienz von Wertpapieremissionen über Internetplattformen, Nr. 8.

Aufderheide, D., Hybridformen in der Internetökonomie - Gegenstand und Methode eines rechtswissenschaftlichen und institutionenökonomischen Forschungsprogramms, Nr. 9.

Grob, H. L., Brocke, J. vom, Hermans, J., Wissensplattformen zur Koordination verteilter Forschungs- und Entwicklungsprozesse – Ergebnisse einer Marktstudie, Nr. 10.

Becker, J., Brelage, C., Falk, T., Thygs, M., Hybrid Information Systems - Position the Web Information Systems Artefact, Nr. 11.

Brocke, J. vom, Hermans, J., Kontextkonstruktion in Wissensmanagementsystemen – Ordnungsrahmen und Ergebnisse einer Marktstudie, Nr. 12.

Holznagel, B., Jungfleisch, C., Die Verwirklichung von Zuschauerrechten im Rundfunk - Regulierungskonzepte zwischen Theorie und Praxis, Nr. 13.

Bröcher, J., Hoffmann, L.-M., Sabel, T., Der Schutzbereich des Markenrechts unter besonderer Berücksichtigung ökonomischer Aspekte, Nr. 14.

Holling, H., Kuhn, J.-T., Freund, P. A., Anforderungsanalysen für Wissensmanagementsysteme: Ein Methodenvergleich, Nr. 15.

Becker, J., Hallek, S., Brelage, C., Fachkonzeptionelle Spezifikation konfigurierbarer Geschäftsprozesse auf Basis von Web Services, Nr. 16.

Brocke, J. vom, Hybridität – Entwicklung eines Konstruktionsprinzips für die Internetökonomie, Nr. 17.

Gutweniger, A., Riemer, K., Potenzialanalyse – Methoden zur Formulierung von E-Business-Strategien, Nr. 18.

Riemer, K., Totz, C., Der Onlinemarketingmix – Maßnahmen zur Umsetzung von Internetstrategien, Nr. 19.

Riemer, K., Web-Design: Konzeptionelle Gestaltung von Internetanwendungen, Nr. 20.

Riemer, K., Müller-Lankenau, C., Web-Evaluation: Einführung in das Internet-Qualitätsmanagement, Nr. 21.

Müller-Lankenau, C., Kipp, A., Steenpaß, J., Kallan, S., Web-Evaluation: Erhebung und Klassifikation von Evaluationsmethoden, Nr. 22.

Müller-Lankenau, C., Terwey, J., Web Assessment Toolkit: Systemdokumentation, Nr. 23.

Müller-Lankenau, C., Terwey, J., Web Assessment Toolkit: Benutzerhandbuch, Nr. 24.

Müller-Lankenau, C., Rensmann, B., Schellhammer, S., Web Assessment Toolkit: Entwicklerleitfaden, Nr. 25.

Gauer, S. S., Evantschitzky, H., Ahlert, D., Kolhatkar, A. A, Marketing innovative Service Solutions with Inter-organizational Service Networks: Opportunities and Threats, Nr. 26.

Holznagel, B., Rosengarten, V., Der Zugang zu Premium-Inhalten insbesondere für Multimedia-Anbieter, Nr. 27.

Zimmerlich, A., David, D., Veddern, M., Übersicht B2B-Marktplätze im Internet Branchenspezifische B2B-Marktplätze - empirische Erhebung, Nr. 28.

Becker, E., Akzeptanz von Internetwahlen und Volksabstimmungen - Ergebnisse der Umfrage zum Wahl-O-Mat in Schleswig-Holstein, Nr. 29.

Totz, C., Potenziale und Herausforderungen der Markenführung im Kontext internetbasierter Interaktionen, Nr. 30.

Holznagel, B., Bonnekoh, M., Auswirkungen der TK-Regulierung auf die Internetmärkte dargestellt am Beispiel von Voice over IP, Nr. 31.

vom Brocke, J., Hermans, J., Anreizsysteme zur Wissensteilung in Netzwerken. Fachkonzeptionelle Modellierung und Prototypische Implementierung für die OpenSource-Plattform HERBIE, Nr. 32.

vom Brocke, J., Altfeld, K., Nutzung von Semantic Web-Technologien für das Management von Wissen in Netzwerken. Konzeption, Modellierung und Implementierung, Nr. 33.

Ahlert, D., Evanschitzky, H., Thesing, M., Zahlungsbereitschaft im Online Handel: Eine empirische Untersuchung mittels der Conjoint Analyse, Nr. 34.

Holling, H., Freund, P. A., Kuhn, J.-T., Webbasierte Evaluation eines Wissensmanagementsystems, Nr. 35.

Trauten, A., Schulz, R. C., IPO Investment Strategies and Pseudo Market Timing, Nr. 36.

Hoffmann, M.-L., Marken und Meinungsfreiheit – Virtuelle Brand Communities auf dem kennzeichenrechtlichen Prüfstand, Nr. 37.

Trauten, A., The perceived benefit of internet-based Commercial Paper issuance in Europe – A survey, Nr. 38.

Ricke, Thorsten, Triple Play – Zugangsansprüche bei vertikalen Verflechtungen, Nr. 39.

Ricke, Thorsten, Neue Dienstekategorien im Zuge der Konvergenz der Medien, Nr. 40.

Müller, Ulf, Utz, Rainer, Aufderheide, Detlef, Meyer, Lena, Rodenhausen, Anselm, Die Zukunft der Internetadressierung: ICANN, DNS und alternative Systeme - kartell- und markenrechtliche Fragen und ihr ökonomischer Hintergrund, Nr. 42.

Holling, Heinz, Freund, Philipp Alexander, Kuhn, Jörg Tobias, Salascheck, Martin, Benutzbarkeit von Software: Wie usable sind Evaluations-Verfahren?, Nr. 41.

Müller, U., Utz, R., Aufderheide, D., Meyer, L., Rodenhausen, A., Die Zukunft der Internetadressierung: ICANN, DNS und alternative Systeme — kartell- und markenrechtliche Fragen und ihr ökonomischer Hintergrund, Nr. 42.

Müller, U., Meyer, L., Unternehmenstransparenz und Geheimwettbewerb im digitalen Umfeld, Nr. 43.

Ahlert, D., Evanschitzky, H., Thesing, M., Kundentypologie in der Multikanalwelt – Ergebnisse einer online- und offline-Befragung, Nr. 44.

Müller, U., Meyer, L., Wettbewerb und Regulierung in der globalen Internetökonomie: Eine rechtsvergleichende Studie zwischen europäischem und US-amerikanischem Recht, Nr. 45.

Becker, E., Bünger, B., Die Rolle des Internets in politischen Willensbildungsprozessen: Ergebnisse einer empirischen Analyse des Internets anlässlich der vorgezogenen Bundestagswahl 2005, Nr. 46.

Berg, C., Döge, B., Pfingsten, A., Internetökonomie im Privatkundenkreditgeschäft deutscher Banken – Theoretische und empirische Beobachtungen, Nr. 47.

Ahlert, D., Heidebur, S., Michaelis, M., Kaufverhaltensrelevante Effekte des Konsumentenvertrauens im Internet - eine vergleichende Analyse von Online-Händlern, Nr. 48.

Schröder, R., Die Neuen Informationstechnologien als Gegenstand der ökonomischen Bildung, Nr. 49.

Trauten, A., Langer, T., Information Production and Bidding in IPOs - An Experimental Analysis of Auctions and Fixed-Price Offerings, Nr. 50.

Grob, H. L., Vossen, G. (Hrsg.), Entwicklungen im Web 2.0 aus technischer, ökonomischer und sozialer Sicht, Nr. 51.